DOCUMENT RESUME

ED 131 929 PS 008 948

TITLE	National Day Care Study First Annual Report, 1974-1975. Volume II: Phase II Design.
INSTITUTION	Abt Associates, Inc. Cambridge, Mass.
SPONS AGENCY	Office of Child Development (DHEW), Washington, D.C. Day Care Services Div.
PUB DATE	Jan 76
CONTRACT	HEW-105-74-1100
NOTE	55p.; For other volumes of this report, see PS 008 947-949
EDRS PRICE	MF-$0.83 HC-$3.50 Plus Postage.
DESCRIPTORS	*Annual Reports; Criteria; *Day Care Programs; *Early Childhood Education; *Educational Policy; *Federal Programs; Government Role; Research Design; Research Methodology; Research Tools; Site Selection
IDENTIFIERS	*National Day Care Study

ABSTRACT

Volume II of the First Annual Report of the National Day Care Study (NDCS) funded by the Office of Child Development presents the research design for Phase II of the study and includes a detailed technical discussion of the study's analytical and methodological issues. The NDCS, being conducted over a period of three years, is designed to answer major policy questions about federally funded day care centers and the children they serve. Volume II is organized into six chapters: (1) Phase II Objectives; (2) Research Design Overview; (3) Discussion of Variables; (4) Research and Analysis Methodology; (5) Testing and Validation of Phase II conclusions; and (6) A Framework for Phase III Planning. (MS)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED131929

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

FIRST ANNUAL REPORT 1974-1975

Volume II
Phase II Design

PS 008948

National Day Care Study

Abt Associates • 55 Wheeler Street • Cambridge, Massachusetts • 617/492-7100

ED131929

National Day Care Study

FIRST ANNUAL REPORT
1974-1975

Volume II
Phase II Design

Prepared for:

Day Care Services Division
Office of Child Development
Office of Human Development
Department of Health, Education
and Welfare
Washington, D.C.

Government Project Officer:
Allen N. Smith

Contract No.: HEW-105-74-1100

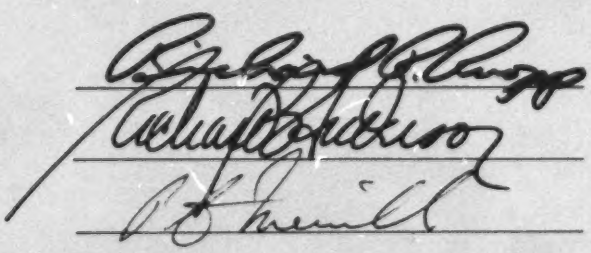
By:

Abt Associates Inc.
Cambridge, Mass. 02138

Study Project Director:
Richard R. Ruopp

PS008948

This report has been reviewed for general publication and meets the quality requirements set by the Research Design Group of Abt Associates. The statements and conclusions contained herein are those of Abt Associates and do not necessarily reflect the views of the sponsoring agency.



Project Director

Technical Reviewer

Management Reviewer

TABLE OF CONTENTS

page

ACKNOWLEDGEMENTS i

PREFACE TO VOLUME II iii

Chapter 1: PHASE II OBJECTIVES 1

Chapter 2: RESEARCH DESIGN OVERVIEW 5

TIMING OF PHASE II DATA COLLECTION 7

THE PHASE II RESEARCH PLAN IN OUTLINE 7

Chapter 3: DISCUSSION OF VARIABLES 11

MAJOR INDEPENDENT VARIABLES 11

The Three Policy Variables 11

Other Center Variables 14

BACKGROUND VARIABLES 15

Parent, Child & Household Variables 15

Site Variables 15

CLASSROOM PROCESS AND CHILD AND PARENT OUTCOME VARIABLES 16

Classroom Process Variables 17

Child Outcome Variables 19

Parent Outcome Variables 20

FINANCIAL OUTCOME VARIABLES 20

Chapter 4: RESEARCH AND ANALYSIS METHODOLOGY 23

GENERAL DISCUSSION 23

DEVELOPMENT OF THE PHASE II QUANTITATIVE MODEL 25

T1 ANALYSIS: GROUNDWORK FOR MODEL EFFORT 25

T1.5 ANALYSIS: CONSTRUCTION OF CLASSROOM PROCESS MODEL 28

T2 ANALYSIS: FINAL CONSTRUCTION OF OUTCOME MODEL 29

ANALYSIS OF DAY CARE COSTS 32

Chapter 5: TESTING AND VALIDATION OF PHASE II CONCLUSIONS 35

SIGNIFICANCE AND POWER ANALYSIS 35

SITE EFFECTS AND GENERALIZABILITY 37

ANALYTIC BIASES 40

Chapter 6: A FRAMEWORK FOR PHASE III PLANNING 43

ACKNOWLEDGEMENTS

The National Day Care Study (NDCS) is being conducted over a three year period by the Day Care Services Division, Office of Child Development (OCD) of the Office of Human Development (OHD), U.S. Department of Health, Education and Welfare (HEW). This first annual report is designed to acquaint the reader with the work of the NDCS from July 1974 to September 1975. Over the past several months the report has undergone a number of revisions as both reviewers and events contributed new perspectives and new information.

Especially important were the views, suggestions and editorial comments of members of the NDCS Advisory Group*: Sally Brown, Virginia Burke, Arthur Emlen, Robert Fein, Mary Howell, Teh-Wei Hu, Jim Levine, Richard Light, Christine Marston, Keith McClellan, Gwen Morgan, Jack Niemeyer, Daniel Ogilvie, Elizabeth Prescott, Mary Rowe, Richard Rowe, and Nancy Travis.

Allen Smith, the OCD government project officer, carefully critiqued each working draft and made substantial contributions. Three of his project consultants, Thomas Achenbach, Robert Brennan, and Victor Cicirelli, made detailed suggestions for revisions in the technical discussion of Volume 2: *Phase II Design.* In addition, Preston Bruce Jr., Director of OCD's Day Care Services Division, made helpful contributions to Volume I.

Finally, in-house review and helpful commentary were made by Richard Anderson and Morris Shepard in their official quality control function and by many of the 33-person staff of the NDCS in Cambridge and at the three study sites.

As with most study reports, this has been a joint effort. Volume I was coordinated by Richard Ruopp, the Foreword and Chapter I were written by him, and Chapter 2 was a joint effort with Craig Coelen. Chapters 3 and 4 are largely Steven Fosburg's work. Robert Goodrich contributed to Volume I, and Jean Layzer and Donna Warner provided editorial assistance. Robert Goodrich is the principal author of Volume 2 with contributions by Marvin Cline, Craig Coelen, and Steven Fosburg, and editorial assistance by Jean Layzer and Richard Ruopp.

Richard R. Ruopp, Project Director
Abt Associates Inc.
Cambridge, Massachusetts
January, 1976

*Background information on members of the National Advisory Committee, OCD Staff and Consultants and Project Staff will be found in Appendix C of Volume I.

PREFACE TO VOLUME II

The National Day Care Study (NDCS) is being conducted over a three-year period by the Day Care Services Division of the Office of Child Development (OCD) of the Office of Human Development, U.S. Department of Health, Education and Welfare (HEW). This first annual report is designed to acquaint the reader with the work of the NDCS from July 1974 to September 1975. It is assumed that the reader of Volume II is familiar with the background material presented in Volume I. An overview of the study taken from that volume is presented below.

Over the past seven years the federal government has become the major institutional consumer of day care for children from six months to 14 years of age. During the 1975 fiscal year alone, close to one-half billion dollars were spent to purchase welfare-related day care for some 550,000 children from both centers and family day homes (35 to 40 percent of available licensed capacity). In the six fiscal years 1970-75, federal, state and local government agencies have spent an estimated total of $2.4 billion on the direct purchase of day care services. About $1.8 billion of this was dispensed by the United States Department of Health, Education and Welfare (HEW), as authorized by Title IV-A of the 1968 amendment to the Social Security Act (SSA)* which provides social service Aid to Families with Dependent Children (AFDC). The 50 states spent an additional $600 million to meet the 25 percent matching requirement, and were responsible for administering all funds.

As a result of its role in day care, the federal government is concerned with the costs and effects on children of day care purchased as a service to welfare-eligible parents. This policy concern has been translated into a set of regulations which apply to those day care programs – homes and centers – serving one or more federally-subsidized children. Since 1968, the Federal Interagency Day Care Requirements (FIDCR) have set standards regulating a wide variety of program characteristics presumed to control the quality of classroom process and thereby both protect children from harm and positively affect their growth and development. For example, the FIDCR specify the minimum number of caregivers that are to be available to each group of children, according to the age of the children. The FIDCR also mandate that at least one caregiver in each group be trained or have demonstrated ability in working with children. These staffing requirements clearly attempt to determine the quality of the day-to-day classroom experience. Equally, these requirements have important cost implications.

The 1968 FIDCR, developed jointly by HEW, the Department of Labor (DOL) and the Office of Economic Opportunity (OEO), were based on expert judgment and the best available evidence. However, no carefully-designed, large-scale studies of day care were available to policy-makers at the time the FIDCR were developed. There were no data illuminating the relationships among day care classroom processes, child outcomes, program costs and the key FIDCR-regulated program characteristics: staff/child ratio, staff professionalism and group size.

The National Day Care Study will assess the costs and the effects of both current and alternative federal day care policies. *The NDCS focuses on the costs and effects on federally-subsidized preschoolers of larger full-day, full-year urban day care center programs.* More specifically, the NDCS is being designed to answer the following major policy questions:

- How is the development of preschool children in federally subsidized day care centers affected by variations in staff/child ratio, staff professionalism, group size and/or other regulatable center characteristics?
- How is the per child cost of center day care affected by variations in staff/child

*Title IV-A was replaced by the Title XX amendment to the SSA in early 1975.

ratio, staff professionalism, group size and/or other regulatable center characteristics?

- How does the cost-effectiveness of federally-subsidized center-based day care change when adjustments are made in staff/child ratio, staff professionalism, group size and/or other regulatable center characteristics?

In December of 1977, the NDCS will complete work on the three major policy questions. At that time, it will present systematic findings to the government on the cost-effectiveness of the current FIDCR and the potential effects of alternative regulations on children, on costs and on the supply of center day care. Study findings will be contained in a series of public reports developed during 1976 and 1977. This report is the first public documentation of the study's background and design.

The First Annual Report is published in three volumes. Volume I, available for general distribution, presents an overview of the NDCS: the background of federal involvement in day care, the day care policy issues being addressed, an overview of the study design and phases, the uses and limits of the study findings and the results of Phase I.

Volume II, which is available upon request, presents the research design for Phase II and includes a detailed technical discussion of the study's analytical and methodolcgical ssues.

Volume III is a compendium of the program, costs, and parent measurement instruments and systems designed for use during Phase II by Abt Associates Inc. (AAI), the research contractor. The battery of instruments selected to measure classroom processes and child outcomes during Phase II is described by Stanford Research Institute (SRI), the testing contractor, in a separate report.

Volume II is organized into six chapters as follows:

Chapter 1: PHASE II OBJECTIVES presents the specific research goals of Phase II as an introduction to the technical materials which follow. Related to the fundamental policy questions discussed in Volume I, these research goals are presented both in general and in detailed form. Each detailed research question is related to specific Phase II research products.

Chapter 2: RESEARCH DESIGN outlines the Phase II design and describes the conceptual model of day care on which it is based. The chapter also presents a schedule of the principal data collection events and a simplified schedule of research tasks.

Chapter 3: DISCUSSION OF VARIABLES specifies the content and construction of the major independent variables that characterize the structure and dynamics of the conceptual model. During Phase II they will be quantified from program, observation, test and cost data so that hypotheses generated from the conceptual model can be tested.

Chapter 4: RESEARCH AND ANALYSIS METHODOLOGY, with Chapter 5, presents the details of Phase II analytic methodology and is divided into two parts: 1) the preliminary development of a quantified model of center day care that relates outcomes to their principal determinants, particularly to the three policy variables – staff/child ratio, professionalism, and group size; and 2) development of an econometric model that relates day care cost to its principal determinants, particularly the three policy variables cited above.

Chapter 5: TESTING AND VALIDATION OF CONCLUSIONS describes the process of testing and establishing the validity (external and internal) of preliminary quantitative results, and includes discussions of analytic biases, attrition effects, and adaptation.

CHAPTER 6: DEVELOPMENT OF PHASE III PLAN discusses how the results of Phase II analysis will be used to develop a research design for Phase III, including choices of centers, sites, experimental manipulations of basic program variables, and choice of dependent measures.

Chapter 1: PHASE II OBJECTIVES

This chapter presents a set of research questions and related research objectives based on the three fundamental policy questions set forth in the Preface.* The research questions translate the NDCS policy concerns into a quantitatively testable format. In later chapters these research questions will be used to discuss testable hypotheses about the dynamics of center-based care and the relationship of center day care processes to day care outcomes. A major product of Phase II research will be an empirically-generated set of hypotheses to be tested by statistical inference in Phase III. In addition, the data collected during Phase II to answer research question 1 below will be incorporated into an early series of policy monographs (see Volume I, Chapter 3) designed to address the issue of FIDCR appropriateness.

Stated in general terms, the central questions for Phase II research are the following:

1. What are the regulatable (and non-regulatable) center characteristics (independent variables), including staff/child ratio, professionalism, and group size, which influence the development of preschool children, satisfaction of parents, and the cost of day care?

2. Which exogenous (background) variables, including family background and child characteristics, mediate the influence of day care, and must therefore be measured?

3. How is the day care classroom process, including the structure and content of interactions among children and caregivers, influenced by center characteristics and exogenous variables? How does the content of classroom process in turn influence child development and satisfaction of parents?

The first of these questions concerns the effect on day care outcomes of the day care center, its program, and its staff. The second question addresses the role of exogenous variables; although these variables are not of fundamental policy importance, they may so influence results that their omission would distort policy-relevant effects. The third question focuses on classroom process. In addition to illuminating the mechanism by which center characteristics are linked to child outcomes, classroom process is important as an intermediate, highly visible outcome measure.

Each of these research questions is further refined below in order to provide greater insight into the goals of Phase II. Question 1 consists of a complex set of definition and measurement issues. It can be expanded to three separate sets of research questions.

1.1 What are the important center characteristics which influence child and parent outcomes and center costs, and how should these variables be defined and measured? What is the composition in the natural day care center world of such complex variables as staff/child ratio, professionalism, and group size? How do these variables differ across centers and over time? How do real world variations in center characteristics affect research definitions? Which characteristics, other than those currently subject to regulation, may be of importance: director qualifications? center auspices? group composition? physical plant structure?

1.2 What is the cost of center day care and how is it related to variations in important center characteristics as defined for analysis?

1.3 How should child development and parent satisfaction be defined and measured? How should quality be defined: as outcomes of service to children? to parents? Should outcome be measured in terms of damage as well

*A detailed discussion of these three policy questions will be found in Chapter 2 of Volume I.

as of development, and what are the significant components of these outcomes? What outcome measures are important to consumers of day care, to advocates, and to the research community?

The next group of questions concerns possible effects of variables external to the center program and structure. These "background" variables may influence outcome as much as or more than the policy-relevant center variables, or may mediate the influences of the policy variables. Research question 2 can be expanded as follows:

2.1 Which family background variables are important influences on child development and parent satisfaction? Do socioeconomic characteristics, parental attitudes, and childrearing practices influence outcome? How should these variables be measured and operationally defined?

2.2 Does a child's previous day care experience influence child outcomes?

2.3 Do contextual factors such as city characteristics influence outcome? Are there differences by site? Must the influence of different factors be considered separately from site to site?

Question 3 considers classroom processes as an intermediary group of variables which link center characteristics to child outcomes. In more detailed form, this question is as follows:

3.1 Which day care classroom processes are most meaningful as measures? How should these process variables be operationally defined?

3.2 Which center characteristics influence classroom process? Do particular combinations of center and program characteristics influence classroom process in special ways?

3.3 Which family background variables influence process?

3.4 What is the influence of a child's previous day care experience on process variables?

3.5 What is the influence of contextual factors, such as city characteristics, on process?

3.6 Which process variables are most likely to influence child outcomes? How does an understanding of these influences and of the influence of center characteristics on process (question 3.2) clarify the influence of center characteristics on child outcomes?

Questions 3.1 through 3.6 indicate that classroom process variables are being treated as intermediate between independent center variables and measures of outcome. They are considered because child and parent outcomes, such as improvements in child test scores or parent satisfaction, may be effects of classroom processes rather than of center characteristics directly. Such explicit consideration of classroom process as an intermediate "domain" between center characteristics and child and parent outcomes will strengthen the conceptual connection between input and outcome domains. External evidence of the connection between process and child development may then be used to strengthen study results.

Each of the three groups of research questions will be answered definitively at the end of Phase III: preliminary results obtained in Phase II will serve to develop testable hypotheses and will be used in constructing the research design for Phase III. The planning of Phase III will, of necessity, rely principally on observed variations in classroom process rather than on children's developmental gains measured through pre-post testing. For the purpose of Phase III design, process variables will be adequate surrogates for test-based measures. Because developmental effects are the direct result of process variations rather than program variations *per se*, variations in child development attributable to the influence of center characteristics must

be preceded by variations in classroom process. Phase III centers will be selected to contain a representative cross section of classroom process.

Each of the research questions will be addressed during Phase II and will lead to a specific research product. Table 1-1 indicates the objective for each question and its projected influence on the design of Phase III and/or selection of Phase III measures. As indicated, answers to research questions 1.1 and 1.2 will be used in the development of four Phase II policy monographs (discussed in Chapter 3 of Volume I).

Table 1-1

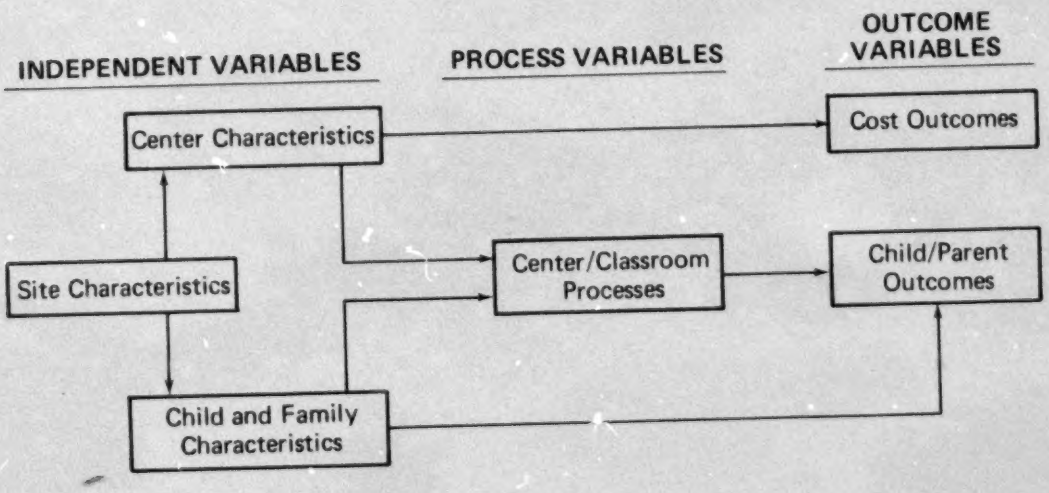
Phase II Research Questions and Uses of Findings

PHASE II RESEARCH QUESTIONS	USES OF PHASE II FINDINGS		
	Primary Influence on Phase III Design	Primary Influence on Phase III Measures	Contribution to Phase II Policy Monographs
1. Center Characteristics			
1.1 Appropriate definition and measure of important center characteristics?	X		X
1.2 Cost of day care?	X		X
1.3 Appropriate definition and measure of child and parent outcomes?		X	
2. Background Variables			
2.1 Appropriate definition and measure of important family variables?	X		
2.2 Influence of child's previous day care experience?		X	
2.3 Influences of site factors?	X		
3. Classroom Process			
3.1 Appropriate definition and measures of important classroom processes?	X	X	
3.2 Influence of center characteristics on process?	X	X	
3.3 Influence of family variables on process?	X	X	
3.4 Influence of child's previous day care on process?	X	X	
3.5 Influences of site factors on process?	X	X	
3.6 Potential influences of process on child outcomes?		X	

An NDCS model of center day care will serve as a common point of reference for issues to be considered in all three phases of the study. Presented in simplest form in Figure 1-1 (it will be expanded in the next chapter), the model reflects our assumption that a group of independent variables describing the site, center, child, and family affects the center and classroom processes, and that those processes in turn impact on children in day care. In addition, characteristics of the child and family directly affect outcome without the intermediary described by center process variables. Finally, center characteristics directly influence center costs. All of the basic study hypotheses will be developed in terms of this simple model.

Figure 1-1

Simplified NDCS Model of Day Care



Chapter 2: RESEARCH DESIGN OVERVIEW

Figure 2-1 is a more detailed version of the model of center day care shown in Chapter 1. The figure lists mnemonic abbreviations for data collection instruments to be used during Phase II, indicating the relation of each to variable groups of the model. Descriptions of the different Phase II instruments are provided in Table 2-1 on page 6.*

Only components relevant to the NDCS are included in this model; that is, the NDCS is focused on full-time care for three-and four-year-olds in study-eligible day care centers. Outcome domains are restricted to those identified in the chart and explained in detail later in this report. Moreover, this research concentrates on the influence of center day care on children and their parents and on the costs of center-based care.

The variables set forth in the model are discussed in Chapter 3. The following paragraphs provide a brief orientation to the variable groups.

Independent variables in this study encompass the policy variables of main interest: staff/child ratio, classroom staff professionalism, and group size. Other center characteristics to be studied include auspices, funding, size, director professionalism; services, curricula and similar descriptors. A set of exogenous independent variables must also be considered. Site and community characteristics may also affect day care process. These include the socioeconomic characteristics of the site and specific structural properties of day care governance and funding at the site level. Outcomes are likely to be influenced by the characteristics of the child and his/her family, such as educational attainment, family income, and family structure; by attitudes and expectations of the family; and particularly by the child's previous experience in day care settings, other early childhood programs, and any other experience likely to influence development. These variables may affect outcomes both directly and indirectly, by mediation of the day care program.

Process variables include both the structural (quantitative) features of classroom interactions, such as the duration and frequency of staff-child contacts, and the psychoeducational (qualitative) content of these interactions. The overall process involves children, parents, and caregivers all together, but it may be viewed separately as child-focused, parent-focused, or caregiver-focused.

Outcome variables represent the impact of day care on children and parents. Final outcome will be measured as positive and negative changes in the child, in terms of social/emotional, cognitive ability and cognitive style development, and in terms of the satisfaction of parents.

As the research is carried out in Phase II and Phase III, the relationships indicated by the model will be quantitatively investigated.* Phase II results will be incomplete and tentative compared to those provided at the end of Phase III, since Phase II is a "natural" study without experimental manipulation or control. Observations and analysis of the carefully selected set of study centers will lead to the formation of a controlled experiment of considerable statistical power for Phase III.

The conceptual model will be converted to a statistically-analyzable form during Phase II. Realistic operational definitions for variables will be developed and statistical relations among variables investigated. The basic mathematical form is to be a regression model, although other forms of the general linear model will also be considered. This process of quantification and hypothesis development is discussed in Chapter 4.

*See Volume I, Appendix B: *Data Collection Systems*, for a full description.

*For a more detailed explanation of the objectives and sequencing of the three phases of the study, refer to Volume I, Chapter 3.

Figure 2-1
NDCS Model of Day Care
with abbreviations of Data Collection Instruments
(cf. Table 2-1)

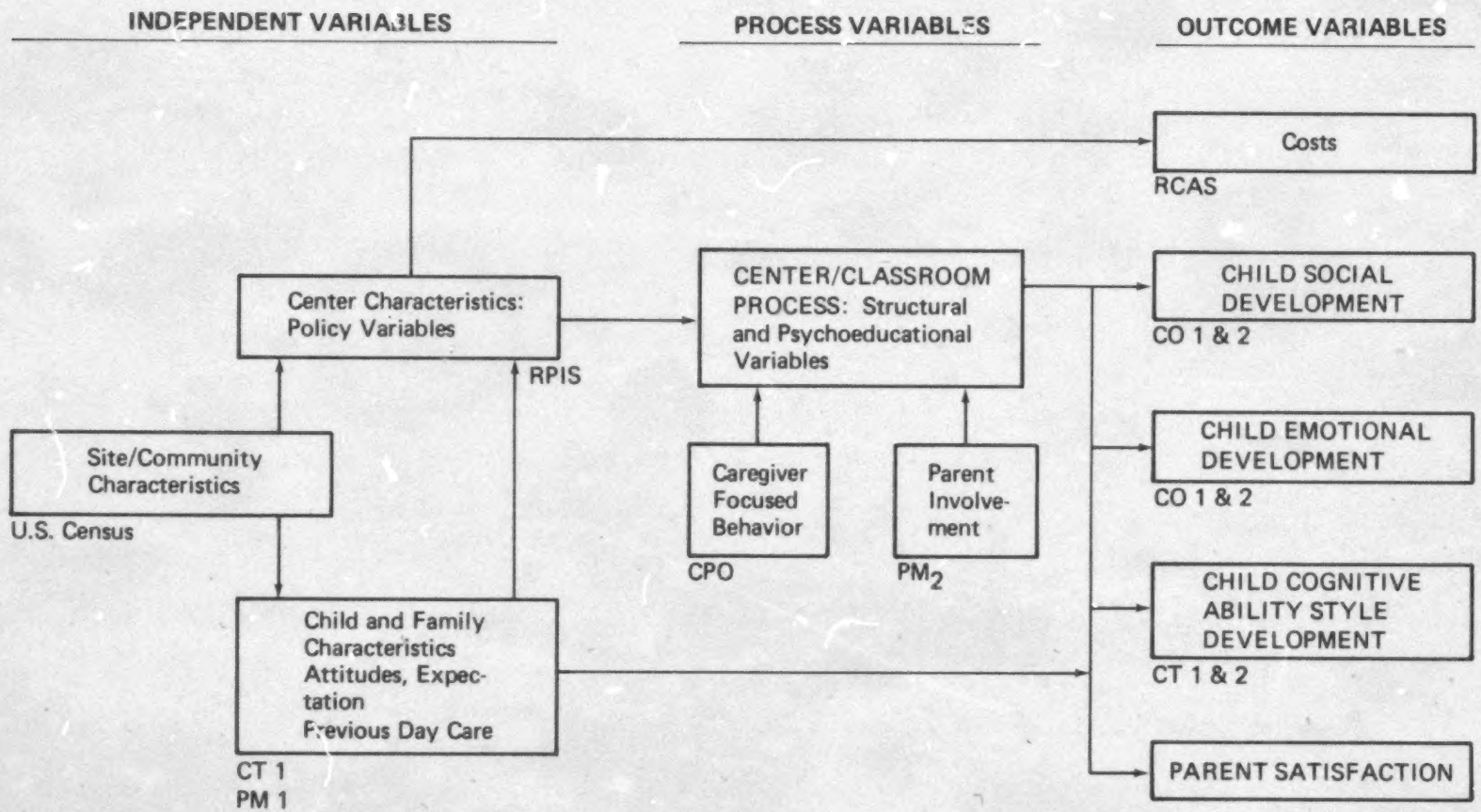


Table 2-1
Data Collection:
Simplified Time Schedule and Instruments

Continuous	**October, 1975 - June, 1976**	
	RPIS	Research Program Information System. Baseline data including center characteristics, policy variables.
	RCAS	Research Cost Accounting System. All center income and expenditures.
T1	**October - November, 1975**	
	PM1	Parent Measures pre-interview, including child, family characteristics, family expectation, values, previous experience of child in day care and other programs.
	CO1	First Child Observations, using revised Prescott instrument.
	CT1	Pre-test SRI test battery.
T1.5	**January, 1976**	
	CPO	Classroom Process Observation. Caregiver-focused behavior, using SRI observation instrument.
T2	**April-May, 1976**	
	PM2	Parent Measures post-interview, yielding satisfaction, involvement data.
	CO2	Second Child Observations.
	CT2	Post-test, SRI test battery.

TIMING OF PHASE II DATA COLLECTION

Three time points in Phase II are central to the analysis plan. Given the symbols T1, T1.5 and T2, they represent the times in fall 1975, winter 1976 and spring 1976 when major blocks of data are collected and available. The times T1 and T2 represent pre- and post-periods for child testing, child observation, and parent measures. Time point T1.5 is the period in which classroom process data will be collected. This schedule reflects the analytic distinction between measures which primarily focus on developmental gains between T1 and T2 and those which concentrate on classroom process as the developmental environment. Table 2-1 provides a simplified outline of this time schedule, including ongoing collections of program and cost data.

THE PHASE II RESEARCH PLAN IN OUTLINE

The NDCS research calendar is best presented in terms of these three data collection points. Phase III specifications of sites, centers, and experimental manipulations are scheduled to

be made after T1.5 and are to be based on analysis of all data collected to that point. The final specification of Phase III measures will follow analysis of T2 data. This section presents the basic outline of the plan, indicating *what* is to be accomplished rather than *how* it is to be accomplished. Methodology is discussed in Chapter 4.

An overview of research organization is presented in Figure 2-2, followed by a discussion of the plan's components. The three groups of Phase II research and analysis tasks are as follows:

A. *Development of Phase II Quantitative Outcome Model (Program and Cost).* Research questions will be addressed through construction of a quantitative model of day care which links final outcome, process, program variables, and background covariables and costs using standard techniques of statistical inference. The quantitative model will be developed as a sequence of milestones, beginning with T1.5 (including process measures but *not* including measures of child development), and then T2 (including a limited analysis of change scores derived from T1 and T2 tests and observations of children). Cost-related research questions will be addressed by construction of a financial model relating day care costs to principal cost determinants, including staff/child ratio, professionalism, and group size. Phase III will continue development and refinement of the quantitative model. Results of preliminary hypothesis testing and effects estimation will be used for Phase III planning during Phase II.

B. *Testing and Validation of Phase II Conclusions.* This task will assess the internal and external validity of the model with respect to analytic bias and generalizability. The results will be used to interpret quantitative conclusions, particularly for Phase III decision-making and for strengthening Phase II and Phase III methodology.

C. *Development of Phase III Research Design.* A basic Phase III experimental design (including choice of sites, centers, and treatment of independent variables) will be developed on the basis of the T1.5 analysis of center/classroom process data coupled with T1 data base analyses. A final, detailed plan, including final specifications of Phase III dependent measures, will be prepared as part of T2 analysis.

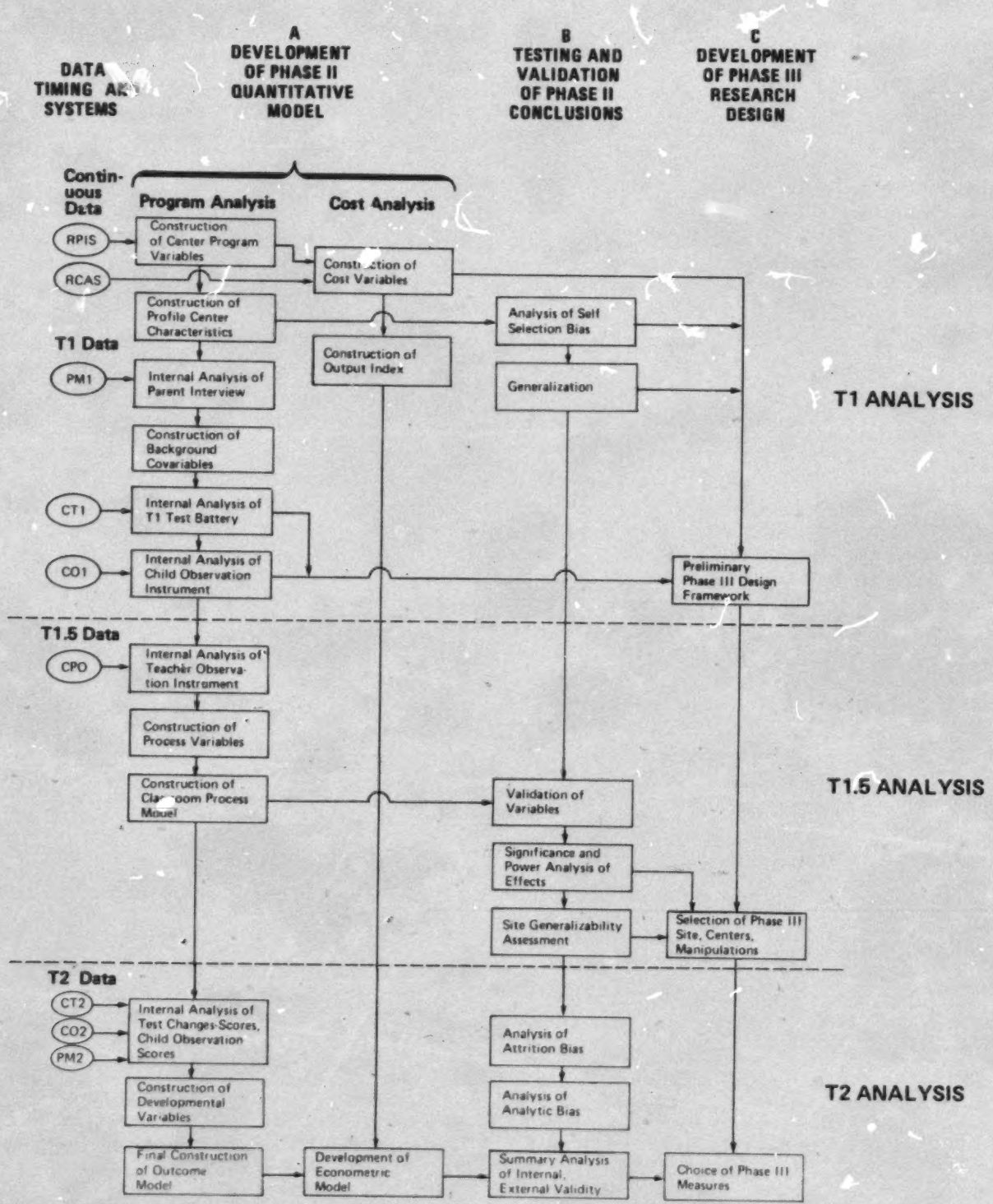
The following discussion relates these tasks to the three major analytic periods that divide Figure 2-2 horizontally.

The *T1 analysis* is principally concerned with developing definitions of variables to be used in later analyses, including center characteristic variables, family background, child and parent pre-measures, cost variables, and economic output indices. Data to be used include: independent program information (RPIS), the parent measures pre-interview (PM1), pretests from the SRI battery (CT1), the first child observation period (CO1), and the cost system (RCAS). Center program variables will be used to structure a profile of center characteristics, which form the independent variables of the analysis. Analyses of the first child observations and the first administration of the test battery will be performed, concentrating on internal psychometric structure. Studies of parent attitudes and generalizability will also begin at this time.

Although the child observation data present one portion of classroom process, they are not sufficiently complete to construct a quantitative model of classroom process. Because many children will not have been exposed to the conditions established by the policy variables long enough to expect any discernible effects, it would be premature to consider the effects of the independent variables on either process or outcome. Conse-

Figure 2-2

Interrelation of Phase II Analytic Tasks



quently, T1 model development will concentrate on the construction and exploration of program variables and background covariables.

T1.5 analysis will be a first step toward assessment of the influence of independent variables by relating them to classroom process measures. This must be preceded by internal analysis of the process data and subsequent construction of process variables, principally from teacher-focused observations but supported by child-focused observation data. The result will be development of the classroom process model.

Construction of a classroom process model is motivated by the assumption that process mediates or fosters development. If this assumption is sound, and if an accurate model of the relationship between process and developmental outcomes can be constructed, it will be possible to use process variables as surrogates for outcome variables in Phase III planning. The sensitivity of different process measures to the independent (policy) variables can be determined directly by observation at T1.5. Through use of the model, plausible hypotheses concerning relationships between the process variables (hence also the policy variables) and developmental outcomes can then be constructed. Such hypotheses will be developed as far as the data allow, since major parameters of the Phase III design must be determined before T2 data are available. In particular, hypotheses based on the model will be useful in choosing sites, centers and experimental variations.

The resulting model will be subjected to a full range of testing and power analysis in order to generate a set of Phase III hypotheses and to provide data for Phase III design decisions. Validation of variable definitions will be carried out, using the process model as a basis. Selection of Phase III design parameters will follow a set of objective decision rules based on the results of Phase II analyses. Site level generalizability will be included as one of the criteria for Phase III choices of sites and centers. The financial model will be completed and the results used in developing the Phase III design.

T2 analysis will focus on the construction of final outcome variables and the Phase III measurement battery. It will make use of T2 child observation and post-test data to extend the analysis from process measures to measures of child development and of parent satisfaction as final outcomes of the process. Analysis will begin with construction of the outcome (dependent) variables derived from an analysis of the internal structure of T2 data. The association of these outcome variables with process variables and policy variables will be investigated to refine Phase III study hypotheses.

Testing and validation activities will include analyses of attrition and analytic biases, as well as a summary analysis of internal and external validity. Analytic conclusions will be presented in the context of these analyses for translation into decisions affecting further research activity, including final verification of Phase III measures.

Chapter 3: DISCUSSION OF VARIABLES

The variables of the NDCS model described in Chapter 2 are developed more fully in this chapter, which is organized by variable domains as follows:

- Major Independent Variables
  - Policy Variables
  - Other Center Variables
- Background Variables
  - Parent, Child and Household Variables
  - Site Variables
- Classroom Process and Child and Parent Outcome Variables
- Financial Outcome Variables

MAJOR INDEPENDENT VARIABLES

The Three Policy Variables

The independent variables include three major policy variables – staff/child ratio, classroom staff professionalism, and group size – and also other potential major determinants of outcome.* At the end of Phase II, each of these variables will have been operationalized and a subset of them chosen for analysis in Phase III. The variables and the approach to this task are described below.

Staff/Child Ratio: Although staff/child ratio is a major cost determinant and program variable, it is often discussed, and even regulated, without being defined operationally. Several different operational definitions will be formulated during Phase II and tested both as potential predictors of costs and child outcome, and also for definitional simplicity and ease of regulation.

A simple definition for staff/child ratio would seem to be the average ratio of children to adults in a class group. Most regulations assume this definition. There are two immediate problems – the child may be with an identifiable core group for only part of the day (in fact, that is usually the case), and children within a group may experience quite different effective ratios in terms of the attention they receive from adults. A child may begin her day in a free play group with a low staffing ratio, join her core group with a higher ratio and then join an activities subgroup. Her nap may be relatively unsupervised and she may rejoin a large group for parent pickup. What is the appropriate ratio? An integrated ratio may overemphasize nap time and entrance and exit periods, at the expense of the core period of the day, when most developmental activities may take place. A core day ratio may overweight a two-hour portion of the day, ignoring six hours of less supervised activities. Within a group, the child may be playing alone, in a small unsupervised informal group or in a supervised small activity group. Thus, the effective contact ratio cannot be accurately portrayed by a single definition of ratio.

On the other hand we may be interested in the staff/child ratio as it is generally described in state and federal regulations. This ratio is generally computed as the ratio of the number of child care staff members to the number of children in a class. Regulations focus on the minimum staff/child ratio allowable rather than the actual amount of adult-child contact during the day, although the amount of contact may be more important in the development of children. Ambiguities appear in the treatment of volunteers and aides who do not regularly appear every day, or work only part of the day.

The 1968 FIDCR regulate staff/child ratio in federally-subsidized centers but do not explain it precisely. A day care center is defined as a facility serving groups of more than 12 children, to which the following regulations apply:*

Day Care Center

a. Three- to four-year. No more than 15 in a group with an adult and sufficient

*For a full discussion of these three variables see both Chapters 2 and 3 of Volume I.

*Regulations cited are quoted from *Federal Interagency Day Care Requirements* DHEW (1968), which is briefly analyzed in Chapter 2, Volume I.

assistants, supplemented by volunteers, so that the total ratio of children to adults is normally not greater than 5 to 1.

b. Four- to six-year. No more than 20 in a group with an adult and sufficient assistants, supplemented by volunteers, so that the total ratio of children to adults is normally not greater than 7 to 1.

c. Six-through 14-year. No more than 25 in a group with an adult and sufficient assistants, supplemented by volunteers, so that the total ratio of children to adults is normally not greater than 10 to 1."

A footnote adds the following information:

> "The adult is directly responsible for supervising the daily program for the children in her group and the work of the assistants and volunteers assigned to her. She also works directly with the children and their parents, giving as much individual attention as possible.
>
> Volunteers may be used to supplement the paid staff responsible for the group. They may include older children who are often highly successful in working with younger children. Caution should be exercised in assigning teenagers supervisory responsibility over their peers."

But these regulations leave the actual computation of a ratio in doubt. For example:

- What does "normally" mean? Should average attendance be used or peak attendance? How do part-time children figure in? What time of day is implied? Do the regulations apply to each hour of each day, or to weekly averages? How, in other words, is the number of children to be computed?
- How is the number of teachers to be computed? Do volunteers count in the ratio? High-school-age assistants? How do part-time volunteers figure in? How do hour-to-hour, day-to-day working schedules enter the computation?

In 1972, OCD drafted a revision of the FIDCR (not implemented). These draft requirements were sensitive to the fact that the actual ratio of caregivers to children does vary through the day and week and could have reduced the ambiguity of the 1968 FIDCR. The 1972 draft described the *computation* of the ratio to be regulated, as follows:

The basic ratio assesses total child hours against total caregiver hours, with the following rules. The ratio is computed by dividing total child hours by total qualified caregiver hours. Volunteers and high-school-age aides do *not* count in the ratio. The computation is at the center level.

There are thus several different formulas by which staff/child ratio may be measured. Each is defensible from some point of view, but if we are interested in effects on children, then we are interested in measures assessing the frequency and duration of contacts with adults throughout the day such as the contact-hour ratio, which compares the total number of daily child hours with the total number of caregiver hours. This definition is also logically extendable to the day care year by estimating average yearly contact ratios on the basis of several time point measurements. We will test all alternative ratios both in terms of their relation to each other and their influence on child outcomes:

- In order to measure head-count ratios that can be implemented by regulation, several different techniques will be employed for weighting full- and part-time children and full- and part-time staff.
- In order to examine process-oriented ratios, several techniques will be used to estimate contact ratios using more detailed program data and also the classroom process data gathered by observation. These data will also allow some assessment of the variation in contacts among children in the same program. Each potential definition must be a plausible one independent of the actual data and particularly of child outcome data.

Not all of these definitions can be directly manipulated by policy regulations. For instance, it would be difficult to mandate the number of caregivers by type of activity and impossible to mandate the amount of caregiver contact time each child must receive each day. There are two important reasons why it is desirable to identify the relationships among the various definitions of staff/child ratio that can be controlled and the actual measured staff/child interactions taking place in classrooms which cannot be controlled directly. First, the most powerful of these relationships will be used in the analysis of the influence of ratio on both process and outcome. Second, to the extent that ratio is influential, these definitions can be used in developing more adequate monitoring procedures for day care licensing and quality control. Therefore, each measure of ratio will be systematically related to all others and will be examined by center characteristics (size, auspices, organization, etc.) and by site.

Professionalism: Professionalism is also a complex measure that may mean different things as a regulatory measure and as a predictor of child outcome. Professionalism is also complicated by the fact that a single measure may not apply equally well to center directors, teachers, assistant teachers, and aides. Regulatory authorities customarily set minimum requirements for professional qualifications for directors, teachers, and aides. Generally these regulations cover several or all of the following areas:

- Chronological age,
- Ability to read and write,
- High school diploma or equivalent,
- B.A., Graduate degrees in general areas,
- B.A., Graduate degrees with concentration in early childhood development, education, or related areas,
- Years of education toward general or early childhood-specific degrees,
- Participation in workshops and special courses,
- Experience in general early childhood care,
- Experience in early childhood group care.

In one state, for example, day care personnel must be "qualified through training and experience to provide good physical care, maintain responsible supervision, and provide meaningful experiences to promote the total development of the children enrolled." This requirement would be unlikely to exclude many potential personnel. Another state regulates the educational level of staff according to size of center. For small centers – 7 to 15 children – the director need be only 18 years old and have a high school diploma. For medium centers – 16 to 39 children – the director also needs one year of experience in family or group day care. For larger centers, the director needs three years of experience in group care, or one year of college and two years of experience, or a B.A. or the equivalent. Teachers must be at least 18 years old and have a high school diploma or the equivalent, while aides need only be 15 years old.

Yet another state evaluates each center's staff with a point system to develop an overall quality rating. The director is assessed by the following formula: seven points for a graduate degree in early childhood education, five points for a B.A. with relevant concentration, one point for each year of post-secondary education up to four points, one point for each two years of work experience with young children up to a maximum of three points, and one to two points for regular attendance at workshops or special courses. Teachers are given fewer points in the same areas, except for graduate work. Aides are evaluated only on experience (two points), workshops (one point) and eighth grade or better education (one point). Such a formula could be used directly in statistical analysis, since the scales are numerical.

These examples show the extreme diversity of requirements across the states, each set of regulations making concrete a state's theoreti-

cal construct of day care professionalism. An important task of the NDCS will be to unify these concepts of professionalism and to operationalize the concept in one or more dimensions.

Operational definitions used in the analysis will depend on the same fundamental measures as do the regulations. Both nominal and interval variables will be considered. For example, years of education and possession of a high school diploma might be considered alternative measures of education. Nominal variables may be useful analytically especially when the threshold is set to equal some important "breakpoint" in the response to the independent variable, but we would usually expect better predictive power from interval scales. The results of analysis of interval variables may suggest good threshold values for regulations.

Before the various definitions of professionalism can be used to measure caregiver behavior in the classroom and/or child outcomes, it is necessary to know how these measures are distributed in our several samples and how various components of the definition of professionalism relate to the process variables. The contribution of various levels of caregiver education, experience, and in-service training to types of classroom processes will be assessed, and findings will be presented by site and by type of center: public or private, profit or non-profit.

Group Size: Group or class size is the number of children assigned to a common physical home base – an area or classroom – under the regular care of the same caregivers. An *intact group* maintains its identity throughout the day while a *non-intact group* may at times merge with others. When children belong to a non-intact group, their group size will be defined as the number in the group during the middle of the day.

Child groups are either *single-age* or *mixed-age*. Study age is defined as the child's age at entry during the current school year – September 1 to August 30. The group *age span* is the span between youngest and oldest children. In mixed-age groups, for instance, there might be a 21-month age span from 2 years 9 months to 4 years 6 months.

Group size has been chosen as an independent variable because it is an easily-measured parameter of center organization. Ambiguities are few but include the effect of absences, changing size over the period of the study, and changing size over the period of the day. Final operational definitions will be constructed to provide high analytic power and to make theoretical and intuitive sense.

Other Center Variables

Other center-level variables may also influence outcomes. The most prominent among these variables are auspices, size, source of funds, director professionalism, curriculum orientation, in-service training, organizational dynamics, parent involvement, and supplementary services. A very few of the likely candidates will be considered as additional independent variables for use in Phase II analyses. If any of these variables prove especially critical, then some reorientation of Phase III may be desirable, but it is likely that such variables need only be considered explicitly in the analysis without affecting the design.

- *Auspices:* The legal organization and economic status of the center. All major forms of auspices are represented in the study population. Included are proprietary, individually-owned and corporately-owned centers; not-for-profit churches; voluntary community agencies; schools; federal, state, and local government centers.
- *Size:* Total center enrollment.
- *Source of Funds:* Predominant funding through parent fees or public money.
- *Director Professionalism:* The training and experience of the director and such other indicators of professionalism that can be readily captured (e.g., director time use).

- *Curriculum Orientation:* Orientation toward custodial or developmental care. The variable requires an adequate operational definition based upon RPIS data data systems.

- *In-Service Training:* The frequency, duration and content of staff training events conducted by the center or other organizations.

- *Organizational Dynamics:* The formal and actual center decision-making structure and processes.

- *Parent Involvement:* The frequency, duration and content of center-organized, individual or group parent activities.

- *Supplementary Services:* The type and quantity of supplementary services to the child (health, nutrition, etc.) or parents (social services, employment, etc.).

BACKGROUND VARIABLES

Parent, Child and Household Variables

Parent, child, and household characteristics are important as uncontrollable independent covariables which may affect outcomes. It is critical to the analysis to take account of family characteristics, attitudes, and history which may be predictive of child outcome. The use and power of such child-level covariables in the analysis are discussed in the next chapter. The following constructs are under current consideration, although others may emerge as equally important after Phase II analyses are complete.

Child's Previous Experience with Day Care. Adaptation to the day care social environment is most rapid during the first few months, and children who have already largely adapted will show less development between T1 and T2. If these children are to be compared to children without previous day care experience, such experience must be taken into account.

Parent Expectations and Attitudes. These covariables will include parents' perceptions of their children's problems, attitudes toward center services, disciplinary practices, school readiness, and a study of parent cultural background.

Parent Perceptions of Children's Characteristics such as persistence, aggression, curiosity, autonomy, and dependency may be useful predictors.

Family Characteristics such as family structure, income, education and occupation may be equally powerful. The most familiar of these variables is the composite variable, socioeconomic status (SES). SES has different operational definitions, all directed toward measurement of a concept of status which is not exhaustively defined or definable, but traditionally includes assessments of economic status, occupational status, and educational status. In the National Day Care Study, SES is being estimated from data gathered during parent interviews or from short telephone inquiries. Preliminary analysis of the data has shown that mother's education and family income have good distribution and are not weakened by excessive missing data. These variables may be combined as a composite measure of SES or may alternatively be retained as separate variables, depending on their explanatory power.

Site Variables

The relationships between site and day care center may systematically influence center process. These possible influences will be considered, to allow us to interpret differences among the three sites.* This section presents some of the theoretical constructs that will be used in an examination of the influence of site on centers.

---

*The possible existence of a "site effect" on parent and child outcomes, with its implications for the generalizability of study results, has emerged as an important issue for Phase II and Phase III. Both Chapter 3 of this volume and Chapter 4 of Volume I have a fuller discussion of this issue.

The following site variables will be considered in interpreting both Phase II and Phase III results:

- *Political Variables:*
  a. Structure of official, political and administrative organizations relevant to day care;
  b. Representation of neighborhoods in city-wide and metropolitan day care organizations;
  c. Role of political parties in day care affairs;
  d. Major civic, voluntary and social organizations active in day care affairs of the site;
  e. Characteristics of site political leaders and other influentials and their involvement in day care.

- *Economic Variables:* To the extent possible a profile of major economic units in the site will be developed by type of industry and by type of occupation, enumerating individuals employed in the categories of professional, laborer, manager, etc. U.S. Census definitions of Occupation and Industry will be used. This will allow us to relate parent data to the broader employed population.

- *Demographic Variables:* Census data will be used to establish a better understanding of the composition of neighborhoods within which study centers operate.

CLASSROOM PROCESS AND CHILD AND PARENT OUTCOME VARIABLES

The complete specification of classroom process and child outcome variables for Phase III cannot be accomplished until all appropriate measures and constructs have been identified and instruments selected and tested, a Phase II analytic task. OCD approved a preliminary group of constructs, associated measures, and an instrument battery for the Phase II pre-test (T1 and T1.5) as recommended by SRI with Abt's concurrence.*

The primary consideration in selecting both process and outcome variables flows from the central question of the National Day Care Study: what are the effects of the policy variables, including staff/child ratio, professionalism, and group size, on both classroom processes and children? Since assessing this relationship is a central goal of the study, it was necessary to identify those classroom and child behaviors which could realistically be expected to reflect the set of experiences provided to children in day care centers. Extreme care was taken to avoid selecting process and outcome measures which would measure behavior in children irrelevant or incidental to these experiences. Selection of such variables could produce the erroneous result of a failure to reject the null hypothesis concerning the true effects of variations on the policy variables under study. Center selection procedures have ensured that variations in the policy-relevant variables will be highly visible. Since Phase III will be an investigation of the effects of manipulating such variables, Phase II must determine systematic relationships among policy, process, and outcome variables to provide a strong foundation for Phase III hypothesis testing.

---

* The instruments selected are:

Observation Instruments
Prescott-SRI (child-focused)
SRI (adult-focused)

Test Battery and Associated Instruments
64 Item Preschool Inventory (PSI) with Hertzig Birch scoring
McCarthy Verbal Memory Test (MVMT)
Matching Familiar Figures (MFFT)
Pupil Observation Checklist (POCL)

A brief discussion of these instruments can be found in Appendix B of Volume 1, First Annual Report. A report by SRI providing a full description of each instrument and field test results will be available in June, 1976.

A second consideration in the selection of variables is that of social relevance. Several different constituencies – parents, caregivers and federal officials – have interests and values that should be reflected in the results of this study. Nearly all groups and individuals involved in the world of child care are concerned that no damage be done to children in publicly-funded institutions. Data demonstrating that enrollment in day care centers regulated by the FIDCR does not result in physical, cognitive, or emotional harm to children will be important to policy-makers, parents, and caregivers, whether or not positive effects of day care may be clearly demonstrated.

Another consideration is that effects of variables should be observable over a period of approximately six months from pre- to post-testing in Phase II. Although the study is not primarily longitudinal, there will be a few children who enter day care as three-year-olds and reappear in Phase III as four-year-olds to provide a small, longitudinal sample that should not be ignored analytically. Measures selected must nevertheless be sensitive to short-term changes for the majority of the study sample of children if they are to provide adequate data in Phase III.

A fourth consideration is the desirability of generating normative data about the impact of various center characteristics. Since this study is directly concerned with the cost of effects, it must trade costs off against the relative value of those effects. Some effects may be important in the statistical sense but may be trivial from a developmental point of view. (A gain of two points on the PS1 may be statistically significant, but of no practical significance to the growth process of children and may not be worth the money or effort to achieve.) The preferred approach is to use instruments to measure variables for which normative data are available so that change in the day care study population can be assessed against change in an independent general population of subjects. This way, the importance of change in developmental terms as well as in statistical terms can be assessed.

Finally, instruments used to assess process and outcome variables must be psychometrically sound. Some of the technical criteria applied are:

1. Testing and observation periods should be short enough so that children will not become exhausted or irritated, and instruments should be attractive enough that test-taking anxieties do not interfere with performance.
2. Tests should be appropriate for both three- and four-year-olds, so that norms and change scores spanning the full two-year range can be established.
3. Tests should have adequate test-retest and inter-rater reliability no lower than 0.6 and preferably higher.
4. There should be an extensive body of knowledge associated with each test's experimental history, including available norms by age and by sex.
5. Instruments should have adequate content validity.

"Adequate" test-retest reliability, inter-rater reliability, and content validity are not rigorously defined quantities. The scarcity of measures for three- and four-year-olds may dictate some flexibility in accepting low values for these measures if the instrument is otherwise superior. In practice, values as low as 0.6 might be accepted for the psychometric reliabilities.

The two major categories of variables discussed in this section are *classroom process* and *child outcome*. They are considered together because classroom process is not only the link between the independent variables (center characteristics) and child outcomes, but also is an important outcome in and of itself.

Classroom Process Variables

Two kinds of process variables, *structural* and *psychoeducational*, are being considered in the NDCS:

1. *Structural Process Variables* describe the organizational and management dynamics of the classroom, including frequencies and durations of interactions between children and caregivers. The variables in this category include:

   a. *Activity Structure* of the class, referring to the balance between caregiver-defined programs (organized instructional periods, group play activities) and child-determined activities (free play or unplanned programs constrained only by physical space and materials available to the children).

   b. *Class Subgroupings.* This variable refers to the size and number of subgroups, their persistence over time, and the extent to which they are spontaneously formed by the children or created by the caregiver.

   c. *Rates of Interaction* between children and caregivers. These involve estimates of rates of interaction within content categories such as:
      - Social interaction /social skill acquisition
      - Intra-personal controls
      - Cognitive/language skill development
      - Physical skill development

These aspects of classroom organization refer less to the psychological content of classroom events than to their structural character. Significant variations in these structural properties are associated with staff/child ratios and the professional preparation of the caregivers (cf. Prescott, 1967). Large centers tend to demand that relatively more attention be paid to management issues than interpersonal relations between caregivers and children, but this dynamic depends on individual caregiver style, physical resources available in the center, the size of classes, and number of caregivers available for managerial tasks.

2. *Psychoeducational Process Variables.* The most pervasive child effects are expected to emerge in the social aspects of behavior, so it is important to measure caregiver behaviors that can be related to variations in social behavior among children, such as:

   a. Statements of rules and constraints, both those imposed by the physical environment and those dictated by social convention.

   b. Discussion and explanation of restrictions

   c. Assertion of arbitrary restrictions

   d. Distinction between motives and acts, feelings and behaviors

   e. Punitive or non-punitive techniques for administering discipline

   f. Rate and mode of interaction with children and with other adults, including openness to physical contact

   g. Classroom management techniques

   h. Provision of materials and classroom structure

Classroom process variables, to a large degree, reflect the independent variables as they are implemented in the classroom. Staff/child ratio, group size, and professionalism foster the emergence of certain interaction patterns, both structurally and psychoeducationally. Certain of the structural process variables, such as the number of child-caregiver interactions per hour, will be very sensitive to policy variables such as the staff/child ratio. It is also important to identify the day care experiences that most influence the emotional, social, and cognitive growth of children. A child in day care experiences many new situations – unfamiliar adults are guiding him/her; he/she must share facilities and equipment with other children; and he/she must also share the caregiver's attention with other children. Psychoeducational process variables will be used to describe the nature and effects of such experiences on the development of children in day care. Child outcomes are the measurable effects of both day care processes and normal development on child behavior.

Again, caregiver behaviors will be considered both as consequences of the policy variables and as antecedents of child outcomes. Analytic plans discussed in the next chapter are designed to accomplish both of these tasks, thus establishing a strong base for generating Phase III hypotheses.

Child Outcome Variables

A short summary of the child outcome variables that have been selected follows. Abt and SRI have jointly arrived at (a) a set of criteria for selection of variables; (b) a justification of each selected variable in terms of these criteria; and (c) operational definitions of the variables which allow them to be linked to specific measurement instruments. These matters are discussed in greater detail in a report to be published by SRI early in June of 1976. As the SRI report will make clear, there are conceptual and measurement issues still to be resolved regarding some of these variables; therefore, the list must be regarded as subject to some revision. However, considerable care has been devoted to selection of the present list, and major changes are not anticipated. The SRI report also discusses a number of additional child outcomes for which assessment might be desirable but has not proved feasible for a variety of reasons.

Three principal groups of child outcome variables are currently being utilized – social/emotional; cognitive/linguistic and physical/motor. As will be apparent, some of the behaviors under investigation relate primarily to the issue of potential harm to children in day care (e.g., incidence of anger and hostility,) and some primarily to positive development (e.g., self-assertive interaction with adults and other children), while still others may reflect harm or normal development depending on the degree and manner in which they are manifested (e.g., dependency). It is for this reason that harm is not treated here as a separate variable to be assessed. In more detail, the three outcome variable groups may be broken down as follows:

1. *Social/Emotional Developmental Variables*
   a. *Dependency:* Behavior reflecting the child's needs for instrumental assistance and/or emotional comfort in problem or stress situations.
   b. *Autonomy:* Behavior reflecting the child's efforts to deal with his world in an independent fashion, e.g., by initiating interaction with others, expressing his views or asserting his rights (such behavior is to be clearly distinguished from the hostile or destructive behavior mentioned below).
   c. *Aggression:* Anger, hostility and negative behavior toward adults, other children or the environment. Though a certain amount of anger is presumed to be healthy and normal in pre-school children under appropriate circumstances, consistent high levels of hostility or destructiveness may be taken as indices of difficulties in the child, the center or both.
   d. *Self-Control:* Behavior reflecting the child's ability to master his impulses.
   e. *Social Involvement:* The degree to which the child interacts with adults or other children, single and in groups. This variable is designed to capture an important effect attributed to group care situations, namely an increase in the child's ability and propensity to deal with others. This variable also reflects the child's willingness to participate, and pleasure in participating, in the activities of the center.
   f. *Prosocial Behavior:* Cooperation, helping, or generosity shown toward other children or adults.
   g. *Compliance/Obedience:* Acceptance of adult judgment and of rules governing behavior in the center.

2. *Linguistic and Cognitive Developmental Variables*
   a. *Cognitive skills and content learning*
      i. *Language skills:* vocabulary, syntax.

ii. *Memory:* especially short-term verbal memory, a specific mnemonic skill relevant for school-related tasks.

iii. *Specific skill and concept learning:* familiarily with everyday concepts, e.g., odor, shape, size, number, time; ability to follow directions and to do simple recognition and reasoning tasks.

b. *Cognitive Style*

i. *Reflectivity:* The tendency or ability to defer immediate response to a task or question in order to test alternative responses mentally, so as to arrive at the best overt response.

ii. *Task persistence:* Ability or willingness to keep at a task despite frustration or temporary inability to perform correctly.

iii. *Generation of ideas:* Facility in inventing novel forms of behavior, e.g., in imaginative play, storytelling, artistic creation, etc.

iv. *Problem solving:* Flexibility and effectiveness in recognizing and overcoming obstacles or difficulties.

v. *Curiosity:* Exploratory behavior; active, self-motivated discovery of new features of the environment.

3. *Physical and Motor Development*

Physical growth of children has not proven sensitive to program variation in other studies and is not expected to relate to the independent variables of the present study — staff/child ratio, professionalism and group size. However, several variables relevant to physical and motor development will be included.

a. *Self-help skills:* The child's ability to cope with his own needs, e.g., dressing, feeding and cleaning himself, etc.

b. *Safety and health:* Accident rates and absenteeism due to sickness will be documented and examined as possible indices of physical harm.

Clearly the above list omits variables which are important in the total development of the child and which might be included in an ideal test battery, e.g., self-concept. As will be made clear in the SRI document, such variables have not been included largely because of the difficulty of finding psychometrically sound measurement instruments appropriate to the age-range and ethnic diversity of the sample, as well as to the practical conditions under which tests must be administered in a national study of the scope of NDCS. However, we are continuing to explore the possibility of locating appropriate measures and adding one or more of these variables to the Phase III measurement battery.

Parent Outcome Variables

Parent outcome variables to be considered include the following:

Parent Involvement. Will be defined operationally by numbers of visits to the center and participation in center activities. This variable is important as an outcome and as a dimension of day care process.

Parent Satisfaction. Another important outcome variable, to be indirectly evaluated from responses to questions about the characteristics of center care, parental use of the resources of the center, and parents' perceptions of the ability of caregivers to deal with children's problems.

Impact on Parent. Parents' ability to deal with problems, employment and so on.

Impact on Child's Behavior. Parents' perception of day care's impact on the child.

FINANCIAL OUTCOME VARIABLES

The primary objective of the financial analysis will be the determination of the effects of policy variables on the per-child cost of day care. Considerable attention will also be paid

to the effects of policy variables on the average compensation received by caregivers, on the percentage distribution of expenditures across budget categories, on fees charged to private-pay day care customers, and on a variety of other financial indicators. The major financial indicators that will be studied during Phase II for sensitivity to variations in the policy variables are:

- *Total resource cost per child:* Value of all resources used in the provision of child care (including in-kind donations) divided by alternative measures of the number of children served. Among these alternative measures are full-day-equivalent child days according to enrollment schedules and full-day-equivalent child days according to attendance records.

- *Net earnings per child:* Total income less total resource cost divided by alternative measures of number of children served. This variable is an important outcome measure because it indicates whether (and to what degree) the center is making a profit, breaking even, or incurring a loss on the average child.

- *Total contributions as a percentage of total resource cost:* The sum of cash and in-kind contributions, foundation grants, allotment from federated fund-raising companies, etc., divided by total resources used in the provision of care. This variable is important as an indicator of the dependence of centers on income other than payments for child care services.

- *Occupancy, equipment, and material costs per child:* This variable should serve as an index of another measure of program enrichment; in-kind donations would be included in the cost measure.

- *Average tuition rate to private-pay customers:* Average fee charged to parents whose child care is not paid for with federal funds. This variable can be used to measure the indirect effects of variations in policy variables on the price of child care to those not federally subsidized.

- *Average compensation per caregiver:* Two measures will be used for this variable, average gross salary and average gross salary plus fringe benefits. Estimation of the effects of policy decisions on these compensation variables is important as an intermediate step in estimating the impact of policy decisions on total child care costs. Compensation variables are also outcome variables in their own right, since part of the benefits provided by the day care industry are the employment and income it creates for its employees.

- *Value of non-administrative professional services per child:* Value of all professional services paid for or received as in-kind donations other than services directly related to center administration (lawyer, accountant, etc.) divided by alternative measures of the number of children served. This variable should serve as an index of one aspect of program enrichment; as such, its sensitivity to variations in policy variables is important.

Chapter 4: RESEARCH AND ANALYSIS METHODOLOGY

This chapter presents study methodology in detail. Discussion covers general consideration of analytic issues and a task-oriented plan for developing a quantitative model of child and cost outcomes.

GENERAL DISCUSSION

A quantitative model of child outcomes and center costs is being constructed so that the Phase II research questions can be addressed and hypotheses for Phase III testing can be developed. Multiple regression will be used to relate center-level independent variables, background covariables, process variables, and both child and cost outcome variables.

A diagram of the model, presented in Figure 4-1, indicates the four levels of analysis (site, center, classroom, child) and the logical progression from independent variables and covariables through process variables to outcomes; and also from independent variables and covariables directly to outcomes. The arrows express the relationships which might be assumed in a particular model. Center characteristics and child/family variables are assumed to produce variation in center process. Child outcomes are directly influenced by process (and thus, indirectly by center characteristics and child/family variables), and directly by client-level variables. Data are being assembled at the child level and preliminary analyses will be conducted at that level. However, many of the important policy questions are at the group, center, or even site level. This analytic interest in the group, center and site does not conflict with the choice of the child as the basic unit of analysis required if best use is to be made of child test data and child/family covariates in regression analysis.

The model will be analyzed principally using the SPSS* software package. The analysis will treat each child as a "case," linking center and group data to each case. All of the children in a group will have the same values for center and group data; all of the children in a city will have that site's identifier variable. In this manner a hierarchical model will be structured as a child-level regression. At the same time group, center, and site level analyses will be possible through aggregation to those levels.

The strongest argument that the child should be the basic unit of analysis is built on expectations about the fractions of variance in outcome which may be accounted for by different factors. A large fraction of the total variance in child outcome will be accounted for by child-level characteristics such as previous experiences, developmental age and family background. When data are aggregated to group or center level, the information contained in child-to-child variations in outcomes and characteristics is lost, but if a substantial part of the total variance in outcome may be accounted for by child-level factors, the analysis of center- and group-level effects is made more powerful. Although this improvement will not be dramatic, it is important.

A similar analysis occurs in the Coleman report,** where a typical finding presents the percentage of variance in verbal ability for white school children attributable to different combinations of factors:

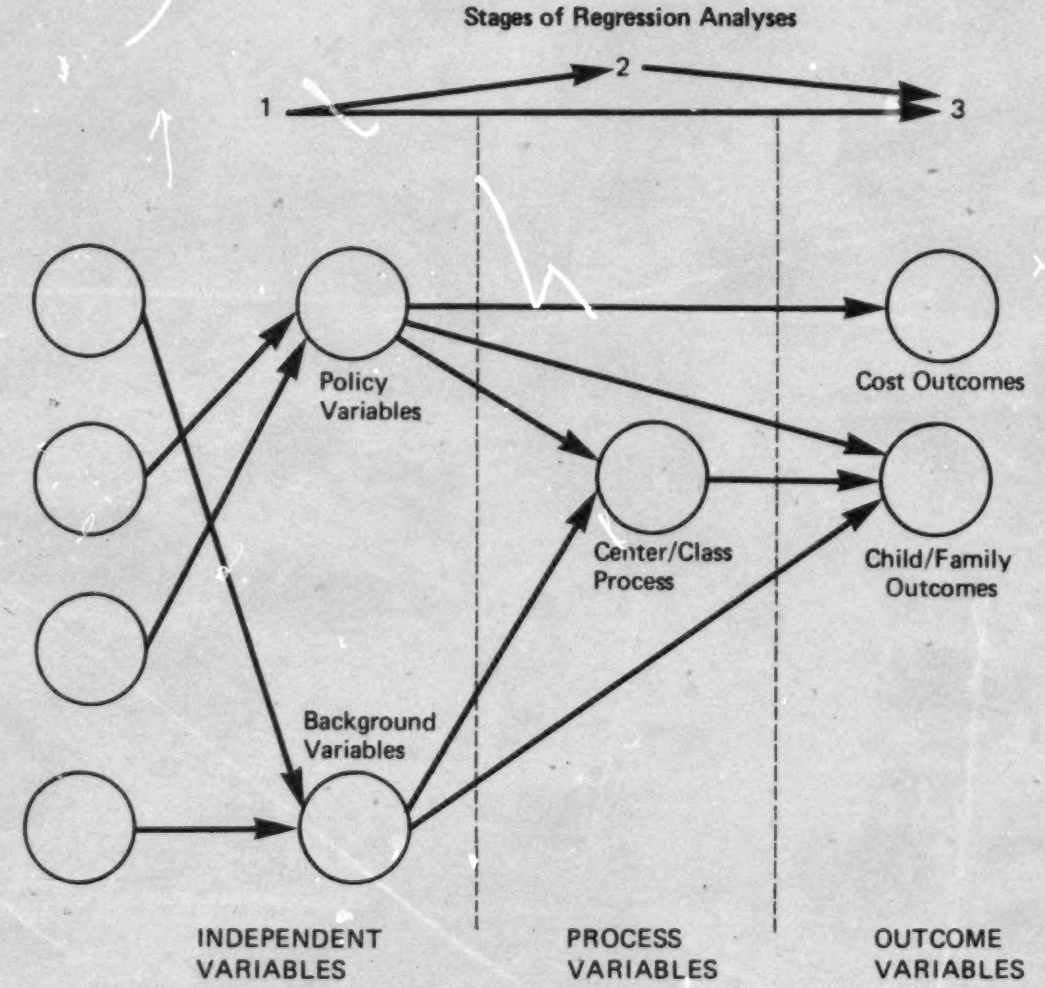
School-to-school differences	13.7 percent
School-to-school plus child background factors	23.0 percent
School-to-school plus background factors plus child's attitudes	36.0 percent

*Norman H. Nie, *et al., Statistical Package for the Social Sciences,* McGraw-Hill, New York, 1970.

**James Coleman, *et al., Equality of Educational Opportunity,* DHEW, Washington, D.C., 1966. Data are adapted from Table 3.221.1.

Figure 4-1
Diagram of Regression Model

NDCS Phases	Number of Cases	Source of Data
II & III	3	Sites
II III	64 32	Centers
II III	128* 64*	Classrooms
II III	1300* 650*	Children/ Families



* Approximate

If the data were aggregated to school level, analyses would compare the 13.7 percent of variance accounted for at this level to the approximately 86 percent which is not. On the other hand, if child-level variables are also included, only 64 percent of the variance will be unaccounted for. This increase in "signal to noise ratio" increases the statistical power of the analysis.*

Many of the variables, including the policy variables, are best defined at the group level, although the group level presents analytic drawbacks that the child-level analysis does not. Groups do not appear sufficiently stable in composition: teachers come and go; children enroll and withdraw; groups are merged and terminated; and new groups are initiated. If these problems are so severe that group level analyses are threatened, the group will be used only to define the policy variables for each child at a given time.

DEVELOPMENT OF THE PHASE II QUANTITATIVE MODEL

Following the schematic of Figure 2-2 in Chapter 2, this section presents the step-by-step Phase II approach to construction of a quantitative model of day care. The effort is divided into three stages that follow the availability of data on independent variables and covariables (T1), classroom process variables (T1.5), the completion of pre-post behavioral and test data (T2), and the collection of nine months of cost data.

T1 Analysis: Groundwork for Model Effort

The objective of T1 is the development of the following quantitatively-defined variables: center characteristic, family background, child and parent pre-test. T1 data include baseline RPIS data, parent interviews (PM1) and first administrations of the test battery (CT1) and child observations (CO1). Analytic tasks to be performed with T1 data include the following:

- *Construction of Center Characteristics*

Approximately 15 center characteristics will be defined and constructed from baseline RPIS data. Variables will include several versions of staff/child ratio, levels of caregiver professionalism, group size, and center organizational data. (Alternative definitions for the staff/child ratio and professionalism were discussed earlier in Chapter 3.) Problems of missing and biased data will be considered in a validity analysis of the definitions (cf. Chapter 5 of this volume and Chapter 4, Volume I.)

Center variables will be tabulated at the center level and used for a more detailed analysis of the distribution of characteristics fully describing the study sample. Center and site selection methods have ensured that staff/child ratio, professionalism, and group size are relatively independent in the sample. Since the number of additional independent dimensions in the sample is unlikely to exceed three or four, a final small set of interpretable center variables will be determined, including the policy variables, which account for most of the variance in center characteristics.

Final outcome of the analyses will be a set of center variables accounting for a substantial portion of the total variance of center characteristics and lying reasonably close to the centroids of corresponding *clusters* of center variables. Thus, each of the variables will be a

*Child-level analyses raise some technical questions. Since correlations must be expected among children in the same class, some of the basic assumptions of child-level statistical estimation are violated. Ordinary least-squares estimates will not be the most efficient estimates, although they will be unbiased. The existence of such an intraclass correlation also invalidates conventional hypothesis testing based on the number of children as the number of degrees of freedom. Greenhouse and Geisser (Samuel W. Greenhouse and Seymour Geisser, "On Methods in the Analysis of Profile Data," *Psychometrica*, Vol. 24, No. 2, p. 95, June, 1959) have shown that F tests involving the computation of "effective degrees of freedom" may be substituted. Their techniques will be used in the analysis.

surrogate for a different variable cluster and will approximate a "factor." These factors will then be used to classify centers by their values for these key variables. The resulting distribution will describe the experimental design as it is implemented. The sample's variation of center characteristics determines the statistical power available to address center-level policy questions. This method of construction of the variable set precludes multicolinearity in the regression.

- *Analysis of Parent Interview*

Parent interviews will be analyzed with three objectives in mind: 1) construction of a profile of parent attitudes, center involvement, background characteristics, and satisfaction with day care service; 2) construction of pre-interview variables to be compared later to results of the post-interview, particularly for assessment of satisfaction related to center program variables; and 3) construction of background covariables for use in the quantitative model. While the first two sub-tasks are straightforward, the third warrants further discussion.

Coleman's data for school-age children lead us to expect that perhaps 25 percent of the total variance in outcome may be accounted for by properly chosen covariables. Two immediate candidates are mother's education and occupational status, both dimensions of SES. The child's previous child care history is very likely to yield a useful covariable, and day care attitudes, child care attitudes, and family structure variables may provide further explanatory power. Since there are no true dependent variable data yet available, construction of covariables must begin not with consideration of their statistical power as predictors, but as useful variables in the child and parent data actually collected. The profile analysis described above is intended to assess these properties. Essentially exploratory, it will result in profiles of age and sex of children, SES characteristics, children's educational history, attitudes and expectations, and similar factors as they relate to center characteristics.

Together with simple statistics, these data will be inspected to eliminate data points which appear unreliable or which vary too little across the sample to be of use.

The next procedure will be correlational and cluster analyses of the data. As with the analysis of baseline data, the objective is to describe the variations in the data, using as few variables as possible. Principal components and cluster analysis will both assist in constructing key variables capturing most of the total variance and representing larger variable clusters. These variable clusters will then be considered as the candidates for child-level covariables in developing the quantitative model (the first analysis of their power as predictors must wait until T1.5 center process variables have been constructed in the spring of 1976).

- *Psychometric Analysis of the T1 Test Battery Data*

The T1 fall test battery included the following instruments administered by SRI:

- 64 Item Preschool Inventory (PSI) with Hertzig-Birch Scoring
- McCarthy Verbal Memory Test (MVMT)
- Matching Familiar Figures Test (MFF)
- Pupil Observation Check List (POCL)

These tests were designed to measure several dimensions of cognitive ability and cognitive style. Psychometric analysis will assess the validity of the data to determine their psychometric properties and to prepare several scales with known statistical and psychological properties for use as variables in general study analyses.

The following specific tasks will be performed:

- Analysis of SRI's report on summer field tests of instruments. Coordinated decisions on Phase II protocols.
- Analysis of a sample of fall protocols for assessment of coding quality, leading to systematic recoding as appropriate.

- Compilation and study of background literature on instruments including psychometric properties, developmental norms, psychological constructs.
- Computation of internal reliabilities, item frequencies, item-total correlations, correlations of subscores and scores among tests.
- Computation of norms for study sample and analysis of variance of scores across centers, sites, sex, and age.
- Assessment of validity and psychometric reliability of scales as outcome variables, with particular attention to expected normal development of scores during the study span.

Recommendations will be developed for operational definitions of cognitive measures to be used as dependent variables. Definitions will require citation of relevant background literature to support their content validity and to reference them to developmental data. In addition the measures must be supported by psychometric analysis internal to the study. Although NDCS data may be used for scale construction, this will not be done in such a way that previously obtained data are invalidated, or that whole new constructs are developed. This analysis will not be complete until T2 data are also available so that the stability of the measures and their developmental properties are known for the NDCS study population.

- *Analysis of Child-Focused Observational Data*

Child-Focused Observational data will be collected at approximately the same times as test battery data, using the SRI-Prescott instrument (modified substantially from an earlier form developed by Elizabeth Prescott). Since this marks the first time that direct preschool child-focused observation has been used in a large-scale study, the potential value of data from this source is so great that the instrument will be further developed as a result of T1 analysis. Most development will be concentrated on sharpening its potential to assess the specific child traits discussed in Chapter 3, such as cooperation, dependency and assertiveness.

Since the instrument is under development and still being tested and refined, it is not yet clear to what extent these two major analytic purposes can be achieved. Actual frequencies of events observed in the field and their psychometric properties are only now being analyzed. The first task is in fact an investigation of these formal analytic properties, somewhat apart from their psychological content. A special computer program has been developed to analyze the statistical properties of T1 SRI-Prescott data. The individual SRI observational codes will first be analyzed, and when the properties of these codes are well known, construction of more complex psychological variables can begin. These may be discussed in terms of the three potential analytic purposes they may serve: 1) analysis of child traits, 2) structural analysis of interactions, and 3) analysis of specific classroom processes:

1) Child traits such as dependency are addressed directly by individual codes, for example, "seeks comfort"; however, the ability to reach conclusions about any pattern of child traits requires both sufficient frequency of the trait behavior within the observation time limit and observer reliability. The use of a trait variable in an analysis of child growth will require high measurement reliability and stability. This reliability and stability is not expected at the individual child level, but it is anticipated that behavior change aggregated to the group and/or center level will be analyzable for sensitivity to variations in the policy variables. Definitions and pre-post analysis will be pushed as far as is feasible. The prime purpose of the work will be, however, instrument development rather than direct analytic usefulness in Phase II.

2) The Prescott instrument data will be analyzed for its usefulness in assessing adult-child and child-child interactions. Each variable, such as frequency of adult-child interactions, may be defined straightforwardly from the observation codes. A limited analysis of time-use by

type of activity seems feasible. The more objective content of these variables makes them easier to measure and to use directly in Phase II analysis.

3) A third analytic purpose is to determine the extent to which the instrument can detect specific processes such as compliance with adult requests. The events which bear upon such a variable may be relatively rare and thus present the greatest difficulty in measurement, and it is probable that T1 data will not allow meaningful definitions of such variables. Because of their psychological importance, however, further instrument development is concentrating on sharpening the system for observing these events, even though preliminary results may indicate difficulties. The payoff for this work will be in Phase III rather than Phase II.

T1.5 Analysis: Construction of Classroom Process Model

Classroom process observations become available at T1.5 using the SRI classroom-focus instrument. The objective of this phase of analysis is to construct classroom process variables, analyze the observation data in terms of these variables, and to integrate the variables with center program variables and background covariables in a preliminary analytic model. The first five research questions about classroom process* may be tested statistically using the model.

The process model will make it possible to consider quantitatively the hypothesis that process measures are causally affected by center variables and the covariables. This preliminary model must be constructed at a group level, since the dependent variables representing classroom process are measured only at that level. Covariable analysis, as mentioned earlier, must wait for the availability of dependent measures at the child level.

Some of the questions to be asked include: How are the process measures related to the independent variables? Which independent variables seem most significant? Do relations appear to be linear or curvilinear? Analysis will begin not with mathematical estimation techniques but with such simple methods as contingency tables and graphical analysis.

A large number of process measures may be considered, with development of the process model used to refine process measures and select those most sensitive to program variation. Approximately six to eight independent variables will be involved, and caution will be exercised to avoid capitalization on chance, which can occur when one of the many possible process measures correlates highly, entirely by chance, with one of the program variables. Thus, the plausibility of relationship must be considered as well as the quantitative data. Phase III allows cross-validation of any suspected relationships with new data.

When the graphical analysis is complete, mathematical regressions will be run between the more sensitive process measures and the independent variables, using the SPSS computer package. The objective here will be to develop a "best" linear model relating each process variable to center-level independent variables. The policy variables — staff/child ratio, professionalism, and group size — will certainly be chosen as three of the regressors, with the remainder selected from three or four possibilities at the center level and three or four to be constructed from child and parent data. The total number of possible regressions is not likely to exceed 2^8, or 256. Virtually all of these may be examined without reliance on such techniques as step-wise regression. The result will be a small assortment of reasonable linear models explaining process.

This analysis is important because it will focus attention on the components of classroom process most sensitive to changes in the program. It may very well lead to improved definitions of classroom process based in part on their correlations with program variables. Although there is not a severe penalty in retaining a

*Questions 3.1-3.5 are listed in Chapter 1 of this volume.

number of process measures at this stage of the analysis, the number must be pared down substantially before the process measures can be used as predictors of final outcomes in order to avoid capitalization on chance. For this reason, the process structure should be described as leanly as possible before analyses of change scores or other final outcomes are undertaken.

T2 Analysis: Final Construction of Outcome Model

The major objective of T2 Analysis is the refinement of the quantitative model relating process measures, independent variables, covariables and developmental measures. This model will extend the group-level process model developed after T1.5, and, as in the development of that preliminary model, graphical analysis will be used to become familiar with the T2 data. This quantitative model of child impacts will be constructed after a set of final outcome variables has been defined, variables dependent on the pre- and post-testing batteries, pre- and post-observations of children, and administration of final parent interviews. The basic design of the model, which will incorporate child and parent covariables, center level program variables, classroom process, and final outcome variables, was discussed at the beginning of this chapter.

Each important problem area in ordinary multiple regression has its graphical counterpart and may be explored graphically, so that problems can be recognized early. Most problems are detectable through examination of residuals after plots have been made, and such graphical residual analysis will be performed as an important step of the process. Mathematical analysis will begin with ordinary multiple regression using the SPSS computer package. Analysis will be univariate in the sense that only one dependent variable may be considered at a time.*

An integral part of the modeling process will be an analysis of residuals for deviations from normality, correlation effects, and heterogeneity of variance. Existence of any of these properties indicates problems with the validity of hypothesis testing and the statistical estimation which must be corrected by rescaling, adding variables or interactions, or considering modifications to the estimation technique. We are particularly concerned about the magnitude of a within-class correlation effect. Even after class-level effects are controlled for in the regression, there are likely to be correlations for test scores and other dependent variables for children in the same classes. The effect can be caused in several ways – observer effect, class-level unreliability, improper modeling, and so on – and will certainly occur to some extent. Process measures in particular involve the whole class and will yield substantial child-to-child correlation. Using the child as the unit of analysis, a within-class correlation of as little as .05 will seriously affect the accuracy of least-squares regression estimators, and will affect the validity of hypothesis development and testing. Consequently, an "effective degrees of freedom" method (Greenhouse and Geisser, *op. cit.)* will be employed to correct sample statistics used in hypothesis tests. Note that this within-class correlation will not bias values for estimates of effects, although their variance will be affected.

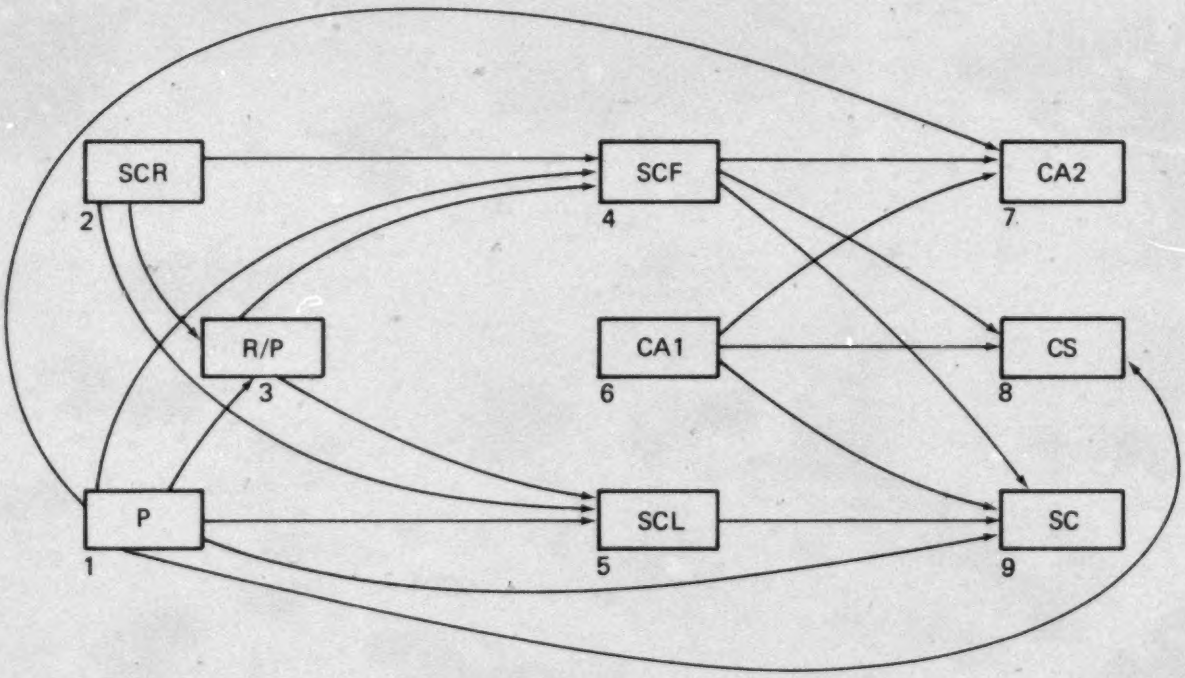
The product of this task will be a model that relates process measures, independent variables, covariables, and developmental measures. If

*Multivariate techniques simultaneously consider several dependent measures. Both ordinary and multivariate regressions consider several independent variables at once. The greater power of multivariate regression derives from its consideration of correlations which may exist among dependent variables. When strong correlations exist, the multivariate model yields estimates for effects which have smaller expected errors and narrower confidence regions than those of the simpler model, which is the principal argument for its use. However, the initial analyses will use ordinary multiple regression because it is considerably less costly in computer time when independent variables are being selected as regressors. It is unlikely that the slight loss of statistical power will have much bearing on the results, but this possibility will be routinely checked.

the potential increases in statistical power appear to warrant it, the model will be extended to full multivariate form, but quite likely immediate needs – planning the experimental design of Phase III – will be well served by the simpler version. (Multivariate regression cannot save a situation which appears hopeless based on ordinary multiple regression analysis.) The resulting quantitative model of day care will be subjected to testing and validation and will summarize the information developed in Phase II.

The logic of the final quantitative model will be presented as a diagram like that of Figure 4-2, which reflects a choice of nine variables and 18 separate paths by which one variable can influence another. The diagram is really a graphic representation of a set of regression equations. It will be conservative in the sense

Figure 4-2
Diagram of Example Regression Relationships Among Variables



1) P: Professionalism
2) SCR: Staff/Child Ratio
3) R/P: P x SCR (Interaction)
4) SCF: Staff/Child Interaction Frequency
5) SCL: Staff/Child Interaction Length
6) CA1: Cognitive Ability (Pre)
7) CA2: Cognitive Ability (Post)
8) CS: Cognitive Style
9) SC: Social Competency

that no reasonable potential effect will be ignored and many unimportant effects are likely to be included so that potentially significant effects can be taken into account in planning Phase III.

Estimates for each relation will be presented as regression coefficients, together with the significance level associated with the related statistical test. A simple form of presentation is indicated in Table 4-1. (While more complicated F- tests or partial F- tests will occasionally be made to test more complex hypotheses, results will be reserved for more technical reports rather than for those for general consumption. Table 4-1 is not directly usable for Phase III planning because it does not contain information on the statistical power of the tests of significance, but Table 4-2 indicates a format for this presentation. Statistical power is a function both of the true size of the effect we wish to detect and of the level used in the significance test (discussed in Chapter 5).* Usefulness of these power tables in planning Phase III will be discussed in Chapter 6 "Framework for Phase III Planning."

*It should be noted that the Phase II model will not exclude potentially important relationships even though they have not been proven significant on the basis of statistical power. Such relationships, even though significant only at .25, may well play an important role in planning Phase III.

Table 4-1

Value of Coefficients

Regression Coefficient I-J of Variable J on Variable I	Level of Significance*	Magnitude of Coefficient
1-4	.06	.1
1-5	.03	.15
1-7	.09	.06
1-8	.29	–
1-9	.48	
2-4	.001	–
2-5	.01	.2
3-4	.16	
3-5	.07	.08
4-7	.16	
4-8	.29	–
4-9	.08	.07
5-7	.31	–
5-8	.06	.1
5-9	.03	.15
6-7	.001	.31
6-8	.02	.18
6-9	.03	.15

*Figures are for purposes of illustration only.

Table 4-2
Power Function Tables
(Shell)

True Size of Coefficient	Statistical Power at Three Levels of Significance		
	At .05	At .10	At .15
.01			
.05			
.10			
.15			
.20			

Analysis of Day Care Costs

Cost analysis will accomplish three specific research objectives. Although there may be some replication of the research in Phase III, definitive answers should be developed in Phase II to the following questions:

- Which regulatable day care characteristics affect the cost of care?
- How should these characteristics, which include staff/child ratio, professionalism and group size, be defined and calculated?
- What is the cost of day care and how does it depend on particular combinations of cost determinants such as the staff/child ratio and professionalism?

Definitions for program characteristics such as staff/child ratio must be the same as those used in assessing child outcomes so that cost-effectiveness analysis can be undertaken at the end of Phase III.

Identifying the major determinants of day care costs will primarily involve development of an econometric model of expenditures from the financial records of day care centers. Various parts of this methodology have been used in previous studies of the cost of child development programs.* The econometric model consists of a set of interconnected equations, each showing the relationship between an important cost variable, whose behavior is to be explained by the model, and a set of explanatory variables. Some of the equations represent tautological accounting relationships (e.g., total operating expense equals the sum of its various components); others will be statistically estimated predictive equations (e.g., an equation explaining the cost of food provided by day care centers in terms of enrollment, number of meals and snacks served on an average day, and an index of the local cost of food).

*Craig Coelen, "Cost-Effectiveness Analysis," *Interim Report V, National Home Start Evaluation,* Abt Associates Inc., October 1974 and Donald Ogilvie, *Estimated Cost of the Federal Day Care Requirements,* Inner City Fund, July 1972.

The model will provide a framework within which the cost implications of policy decisions can be evaluated. Among the explanatory variables in the model will be the important policy variables — staff/child ratio, caregiver professionalism, and group size — as well as the number of federally-subsidized children and the amount paid per child. Also among the explanatory variables will be such exogenous factors as local prices for labor and other input, local population density, local demand for care, and the legal organization of the center. Among the variables whose behavior the model will explain will be cost of care per child, fees charged to private-pay clients, and the centers' net earnings. Once the model has been specified and the parameters in the stochastic equations estimated, it will be possible to simulate the effects of independent variables on the dependent economic variables. Simulations can be performed for alternate sets of values for the exogenous variables to identify differentials in the cost implications of policy decisions from one set of local circumstances to another and for different types of centers.

During Phase II, the forecast/simulation model will be used to identify the policy variables which, when allowed to fluctuate over a relevant range, have the greatest effect on the cost of center-based day care in metropolitan areas. Published data and information from the National Day Care Supply Study's telephone survey will be used to determine the relative weight given to local demand for care, input prices and to different types of centers in aggregating effects to form national forecasts.

Important to Phase II research will be a careful study of the effect of changes in staff professionalism on the cost of the caregiver's time. A number of recent economic studies of the determinants of intraoccupational wage differentials provide theoretical and empirical background. The most widely used model explains the salary paid by the firm (day care center) for a particular type of worker in terms of wages in competing occupations (secretary, aide in public schools, etc.), an indicator of the importance of the particular occupation (percentage of local labor force in the occupation), and a set of quality variables. In the case of day care center staff, quality variables will include years of previous child care experience, general educational attainment, and the number of special courses completed.

Timing of these tasks during Phase II depends on data flow from the research sites. The first important analysis task will be an examination of the relationship between the wage rates paid to caregivers and their qualifications (experience and educational attainment). Data on salary scales and on qualifications from staff background questionnaires will be available early enough that statistical analysis can begin during January 1975, with results to be presented in Interim Report III.

The second major analysis task will involve determination of the relationship between staff/child ratios and the cost of day care per child. A mixed statistical/accounting model of center expenditures will be developed for this purpose. The data necessary to construct this model will arrive from participating centers on a monthly basis, but with a reporting lag. (Data for the month of October, for example, may not be complete and available for analysis until spring of 1976.) The preliminary version of the expenditure/cost model will have to be available for Phase III planning by the beginning of May 1976. On the basis of the expected lag in data availability, the model must be developed with five-month data from October 1, 1975 to February 29, 1976.

A second version of the expenditure/cost model will be available for the Final Phase II Report in early September 1976 based on eight-month data which (unlike the five-month data referred to above) has been converted from a cash recording system to an accrual recording system.*

*This adjustment smooths out large one-time-per-year expenditures (e.g., insurance premiums or taxes) over a 12-month period. Without this adjustment, data on expenditures may overstate or understate the cost of operations during a part of the year.

Chapter 5: TESTING AND VALIDATION OF PHASE II CONCLUSIONS

Quantitative results of the study will be subjected to a technical examination of their usefulness. Because the results of Phase II will be used principally to design Phase III, validity assessment will concentrate on potential effects on Phase III design. Nevertheless, potential problems in external validity must be detected early enough to allow appropriate changes in the Phase III design. Overall assessment areas include the following:

- Significance and Power Analysis
- Site Effects and Generalizability
- Analytic Bias
  - Attrition
  - Self-Selection
  - Analytic Methods
    - Measurement
    - Simultaneous Equations

SIGNIFICANCE AND POWER ANALYSIS

The quantitative models under consideration will have been subjected to a great deal of significance testing during their development, a familiar part of the analytic process. In this case, however, the models will be subjected to a further round of significance testing combined with power analysis to provide more information about the internal usefulness of the findings for planning Phase III. While a researcher might be extremely reluctant to publish a finding statistically significant at the .25 level, it would be irresponsible to ignore its potential use in the Phase III design. Because the concept of power is not a familiar one, this section will explain the application of power theory to the NDCS.

If the independent variables do affect outcomes, how likely is the National Day Care Study to detect these effects? This depends on the size of these effects, and the study will do much better at detecting large effects than small ones. The theory of statistical power treats this problem directly, making it possible to assess, for a given experimental design, the probability of detecting effects of specified sizes. This probability is expressed as the power function. Ideally, the power function would equal one whenever there is actually an effect, and zero otherwise, and would correspond to an experiment which would always produce the correct results. If the null hypothesis were correct, the experiment would say so. If not, the experiment would detect the deviation and reject the hypothesis.

It is regrettable that statistical power is such a complicated technical issue, because it is extremely important to this project. Before data collection, we can assess quantitatively what we intuitively think of as a power of the experimental design. After data are collected and analyzed, some relations will be labeled as *significant* – at some level – and others will be labeled as *not proven significant.* These latter cases, where the null hypothesis is not disproved, may arise either because no effect exists or because the power of the test was not sufficient to detect the effect.

We cannot discard variables from the Phase III design simply because they are not proven significant at some level. Significance tests are usually very conservative, and effects which are practically significant may fail the significance test through pure chance. The theory of power deals with the probability of detecting statistical significance, given the truth of an alternative hypothesis.

Normally, the probability of detecting alternatives which are quite similar to the null hypothesis is very low. Thus, for example, we might find that the probability of detecting a small effect was .1, but for a larger effect, the probability was .5. For practical purposes, if an independent variable produces a very small effect, we may be willing to consider this no effect at all. If, for example, the PSI scores change by only two points we may feel that this is too small an effect to be operationally meaningful. Thus, if the probability of detecting this effect was only .1, it would not be of concern. The power need only be high for effects which we have determined are important.

The analysis to be presented follows that of Jacob Cohen in his book *Statistical Power Analysis for the Behavioral Sciences.* Rather than a pure regression model, Cohen uses ANOVA and ANCOVA as model analyses. This represents a simplification which will not change basic conclusions. The true size of the effect is measured by Cohen's "effect size index" f, which is calculated as $\sigma m/\sigma$ where σm is the standard deviation of means across treatments and σ represents the within-treatment standard deviations. Engineers might refer to f as the signal-to-noise ratio. Multiple R^2 is related to f by the formula

$$R^2 = \frac{f^2}{1 + f^2}$$

For the purposes of this discussion, effects have been classified as small, medium, and large. These definitions relate to f and R^2 as follows:

	f	R^2
Small	.10	.01
Medium	.25	.06
Large	.40	.14

The definitions of these quantities make sense, since in psychology and related fields, effects with R^2 of .2 or more are rare and effects with $R^2 < .01$ are uninteresting (in these terms, Coleman's effects would be classified as medium to large). In center-based day care, background variables of children and families are likely to have large effects and center program variables small-to-medium effects.

Assume for the moment that there are no background covariables available and that an analysis of the main effects of the three defining variables can be carried out by means of an ANOVA. There are two common choices for the significance level for testing the null hypotheses, 5 percent and 10 percent. Statistical power will be estimated for both choices. For 64 observations, the probability of detecting small, medium, and large effects is as follows:

	Small	Medium	Large
5% Level	.12	.50	.88
10% Level	.20	.63	.93

This table illustrates one of the major compromises in statistical analysis. The more stringent the significance level, the poorer the power. Thus, at 5 percent there is only a 12 percent chance of detecting a small effect, whereas at the 10 percent level, there is a 20 percent chance. In both cases, however, the power of detecting small effects is meager. For medium-sized effects the power increases quite substantially, and for large effects, the chance of detecting these effects is about 90 percent. In both cases, the chances are better than 50 percent of detecting an effect. The greatest gain in power between the two significance levels is for medium-sized effects.

Power increases with sample size. For example, an analysis of 128 classrooms would result in the following probabilities that effects would be detected. Increases in power are especially substantial for medium effects. Large effects are almost certain to be observed.

	Small	Medium	Large
5% Level	.2	.81	.99+
10% Level	.31	.89	.99+

This suggests that the power of Phase II design is not sufficient to make detection of small effects highly probable, while medium or large effects are likely to surface. When a small effect is considered to be large enough to be of potential importance to policy, an extremely conservative attitude should be adopted toward "null findings," where effects do not appear significant at the .05 or .10 level. Phase II has sufficient power for efficient design of Phase III.

Gains in power can be achieved by using background information on the children. Child-level data, including both background information and testing results such as PSI scores, can be used in the regression analysis.

The amount of gain in power achieved by using the covariables depends on the correlation coefficient r , which relates the covariable to the dependent outcome variables. Table 5-1 exhibits the effects of covariation resulting in values for r of .3, .5, and .7. In the case of the Coleman data, about 25 percent of total variance in achievement scores could be accounted for by such background and attitudinal variables.

If these results are replicated here, an r of .5 might be expected. Comparison of the charts shows that the use of covariables would increase the probability of observing a null effect from .30 to .38, assuming a 10 percent significance level and a sample size of 128. This increase, which was very conservatively estimated, is far from dramatic but is large enough to justify the use of covariables.

SITE EFFECTS AND GENERALIZABILITY

Generalizability is relevant in Phase II only as it relates to the proper design of Phase III, since NDCS research was not intended to answer the major policy questions until the conclusion of Phase III. Generalization consists of extending conclusions based on the study domain of sites, centers, and independent variables to a larger domain of interest, in this case all American cities and eligible

Table 5-1

Probability of Effect Detection

5% significance level (n = 64)

Coefficient	Small	Medium	Large
r = .3	.13	.55	.91
r = .5	.16	.64	.95
r = .7	.20	.80	.99+

10% significance level (n = 64)

	Small	Medium	Large
r = .3	.22	.68	.95
r = .5	.26	.75	.97
r = .7	.31	.88	.99+

10% significance level (n = 128)

	Small	Medium	Large
r = .3	.33	.9	.99+
r = .5	.38	.93	.99+
r = .7	.49	.99	.99+

centers and a wide range of regulatable program variables. The NDCS domain consists of three study sites and 64 study centers exhibiting a wide range of independent variables (cf. Volume I, Appendix A: "Site and Center Selection").

The chain through which findings are to be generalized is a complex one, depending on internal validity of the analysis and the relationships among study units and the units they represent in the external world. The establishment of internal validity has been treated in several sections of this report; the discussion that follows is on potential "site effects," an issue extremely important to geographical generalizability.

A site effect is a measurable difference in day care process or outcomes attributable to underlying differences among the sites and detectable as heterogeneity of regressions among sites. The existence of a real site effect, suggesting significant local differences in appropriate day care program characteristics, would call into question the advisability of uniform day care regulations imposed on every geographical region. A site effect is also of importance analytically as a factor which may complicate estimation of center program effects.

Phase II statistical data will not be sufficiently powerful to allow attribution of site effects to particular socioeconomic variables such as city size. Since there are only three sites in the study, potential explanatory variables are highly confounded. Every effort is being made to catalog potential contributors to site effects during Phase II for more detailed study in Phase III.

During Phase II site identifiers will be employed in statistical analyses. When all interactions are considered, this approach is equivalent to making separate regressions in three sites. The question, "Is there a site effect?" can be rephrased as, "Are the regressions the same?" Two principal kinds of site effects may be observed, with quite different implications.

Consider the problem as one in analysis of heterogeneity of regression. Using three separate regressions, the models

$$Y_i = \alpha_1 + X_i B_1 + e_i$$

$$Y_i = \alpha_2 + X_i B_2 + e_i$$

$$Y_i = \alpha_3 + X_i B_3 + e_i$$

are fitted to the data from the three sites. If the α's are significantly different site-to-site, then the dependent variable Y is different in the three sites even in the absence of the effects of center program. For instance, average PSI change scores may differ significantly from site to site, independently of program variations within site. These site effects are not policy relevant because they do not imply that the policy variables have different responses to programs among the sites, but simply that average developmental rates may differ among sites.

The more important form of site effect may be recognized as significant differences in the B's. For instance, variations in staff/child ratio might interact with the site variable, so that overall effects of the staff/child ratio differ from site to site. Separate analyses at different sites would result in different conclusions, and regression surfaces would have different slopes. If the slopes are significantly different, we have heterogeneity of regression. Tests for heterogeneity may involve one, several, or all regression parameters. The most interesting test might involve only one B-weight corresponding to the effect of staff/child ratio. But any of the possible ways in which the regressions can differ may be tested by the routine F-test for equality of sets of coefficients from site to site. The test statistic may also be used to construct a confidence interval around estimates of differences among coefficients.

Since the test for a policy relevant site effect depends on statistical comparison of different regressions, a significant effect can exist only if significant regressions exist. If Phase II

should lack the power to demonstrate effects of staff/child ratio and of profesionalism, there is no way that significant site differences in these effects can be demonstrated. Demonstration of a significant site effect requires substantially more statistical power than demonstration of a policy variable effect, and tests for the site effect may fail because no effect exists or because of insufficient power.

Tests for site effect do not explore potential causes for the effect but merely its existence or nonexistence. While a "true" site effect may be caused by differences among the sites, independent of center characteristics, a "false" site effect may result from confounding of center characteristics with site. We know that such confounding exists. For instance:

- *Education level* of caregivers (and of the general population) is higher in Seattle than in Atlanta or Detroit.
- *Demographic characteristics* for the children are different among the three sites.
- *Economic class mixing* is more common in Seattle and Detroit than in Atlanta.
- *Patterns of sponsorship* are different among the three sites.

Any one of these differences may cause an effect which might be mistaken as a true site effect. While these variables may be "partialled out," the fact that they are confounded with site means that a true site effect might be removed as well. It may not be possible to distinguish among these and true site effects unless we see strong variations of outcome within, as well as across, sites.

The number of different tests for site effects is the same as the number of different ways that heterogeneity of regression may be tested. Each test is constructed as an F test for coefficients of dummy variables representing identity. Table 5-2 enumerates these tests for a single independent variable; for several

Table 5-2

Site Effect Null Hypotheses For Regression Coefficients

$a_1 = a_2$	
$a_1 = a_3$	
$a_2 = a_3$	
$a_1 = a_2$	$B_1 = B_2$
$a_1 = a_2$	$B_1 = B_3$
$a_1 = a_2$	$B_1 = B_2 = B_3$
$a_1 = a_2$	$B_1 = B_2$
$a_1 = a_3$	$B_1 = B_3$
$a_1 = a_2$	$B_1 = B_3$
$a_1 = a_3$	$B_1 = B_2$
$a_1 = a_2$	$B_1 = B_2 = B_3$
$a_1 = a_3$	$B_1 = B_2 = B_3$
$a_1 = a_2 = a_3$	$B_1 = B_2$
$a_1 = a_2 = a_3$	$B_2 = B_3$
$a_1 = a_2 = a_3$	$B_1 = B_2 = B_3$

different independent variables, many more combination tests are possible including the well-known Chow test.* The ability to generalize to other cities is illuminated by results of such hypothesis testing, but such results should be understood in the context of statistical power.

Consider the implications of various findings. Suppose that B_1 (Atlanta) $\neq B_2$ (Detroit) at a significance level of .10, demonstrating a probably site effect. In that case, one would be precluded from generalizing B_1 or B_2 to all cities but might be tempted to generalize B_1 to the southern cluster and B_2 to the northern. Since only one case is available per cluster, this generalization cannot be justified.

*G.C. Chow, "Tests of Equality between Sets of Coefficients in Two Linear Regressions," *Econometrica*, Vol. 28, July 1960.

If the effects of the socioeconomic variables were estimable across the study sites, this knowledge might be used to construct estimates of the variance within clusters. One would assume that variance was due to these effects and would use the results of site variable regression. Since the previous section showed that these effects were *not* estimable, within-cluster variance may not be estimated in this manner.

On the other hand, suppose that $B_1 = B_2$ with considerable statistical power. In this case, there is evidence that the site-to-site variances of effects are small. Since the two sites, say Atlanta and Detroit, were selected in order to highlight socioeconomic differences, the evidence is stronger than it would be were the two cities socioeconomically similar, although it is difficult to make this a quantitative argument. If $B_1 = B_2 = B_3$, the evidence is stronger still, and the generalization that B's are nearly equal for all sites is reasonable, particularly when there are other independent data to support the conclusion.

Generalizability, then, is supported when null hypotheses involving the quality of certain regression coefficients are *not* rejected. The strength of the evidence on generalizability depends directly on the statistical power of the tests used, power that can be estimated using the tables presented earlier and following the same conventions regarding small, medium, and large effects. If the effect of a program variable is small, there would be little interest in adjusting it from site to site, because the site effect would be of little importance in a Phase III design. If an effect is medium, then small site effects would be of interest; if the effect is large, then medium site effects would require explicit consideration in Phase III.

Thus, we may construct the following contingencies, based upon 125 degrees of freedom (there are 126 target classrooms in the NDCS as of this writing), as in a group level analysis. Each statement relates the outcome of statistical effects test to the real situation. If the program effect is small in size, the probability of detecting it at a significance level of .10 is about .30, and detection of a site effect is very unlikely. If the program effect is medium in size, there is about a .9 chance of detecting it and a .3 chance of detecting small site effects. If the program effect is large in size, it is practically certain to be statistically significant, and there is a .9 chance that a medium site effect will also be significant. While these statements are made as though we knew independently how large the program effect is, in reality we will see only the results of the statistical test and will have to infer the size of the program effect.

Using the information in the preceding paragraph we can examine the situation from the point of view of reasonable inferences. If statistical tests do not indicate a program effect, then it is relatively certain that the real effect is at least small, and Phase III may be designed without great concern for a site effect. The results, if replicated, would indicate that a substantial effect is unlikely to exist anywhere. If a significantly large program effect is detected, but a significant site effect is not detected, reasonable generalizability to other cities would also be expected. If a medium site effect were detected, on the other hand, generalizability would be threatened and Phase III should be designed with the supposition that a site effect exists.

The most ambiguous results are obtained when a medium program effect is observed. If a site effect is detected, then it is probably real and should be considered. If a site effect is not detected, however, it is not disproved because of low statistical power. The direction that Phase III design should take would require assessment of additional information.

ANALYTIC BIASES

Several forms of statistical bias are likely to be confronted during the course of the National Day Care Study. This section defines the nature and consequences of bias, identifies the circumstances which may give rise to it, and suggests some remedial actions.

Bias is the divergence between the true value of a parameter and the expected value for a particular estimation formula and a particular sampling procedure. If θ is used to denote the true value of the parameter and E (θ) is used to symbolize the expected value obtained from a particular estimation method, then bias is defined as: E (θ) $- \theta$. Biased estimation methods may yield unreliable numerical estimates and tend to distort the results of hypothesis tests. Some bias may be acceptable if the estimator is consistent, so that bias becomes very small for large samples, but inconsistent estimators are nearly always troublesome.

Four potential sources of bias will be considered explicitly in the Phase II analyses. *Attrition bias* and *self-selection bias* are possible consequences of the composition of the research sample. *Measurement bias* and *simultaneous-equations bias* may result from our choice of estimation method.

Attrition is the loss of children in the study, especially of children for T2 testing and observation, who *were* present at T1. They may have dropped out of the study, may refuse participation, or may be absent on the testing date. If the excluded children are different from retained children in some substantial systematic way, the analysis that depends on pre-post data will exclude these children and will be based on a biased sample. *Self selection* occurs when the selection process is not random but depends on the willingness of the subject. Since willingness may depend on the characteristics of the subject, the sample may be distorted and analysis affected.

The approach of the study to attrition bias and self-selection bias should be sequential, a primary task being to measure the degree to which attrition and self selection have occurred. If the occurrence is large (more than 5 percent of the sample) or concentrated in one or a few centers or in certain types of centers, the next step should be a comparison of the characteristics of the subjects remaining in the sample with the characteristics of those who declined to participate initially or who have since dropped out.

Note that not every difference between the two groups is necessarily a source of bias. The difference must be such that efforts to estimate relationships between outcomes and predictive factors is distorted by the composition of the remaining sample. If the group made up of those who declined to participate initially or who thereafter dropped out is large and in any systematic way different from the group of remaining subjects, the Office of Child Development will be advised, and a joint decision will be made on adjusting the Phase III design to eliminate possible distortion of statistical results.

Measurement bias will arise during Phase II whenever the data used for explanatory variables in regression equations are contaminated by measurement error. One source of measurement bias is the use of pre-test scores in regression equations to explain post-test scores. The pre-test instrument is of less than perfect reliability, so the figure entered in the regression does not equal the true score. This may be shown to bias the estimate of the regression of the post-test score or the pre-test by a factor equal to the test-retest reliability. This factor, if it is known from other data, may be used to remove the bias.

The bias just described occurs even when the error in measuring the pre-test score is uncorrelated with the true pre-test score, and the situation is even worse when these corelations exist. But there is a corrective approach in this case, even when better data cannot be obtained. The remedy is to correct the standard estimation formula for multiple regression to reduce the sensitivity-to-measurement error. An estimation technique known as the instrumental variable (IV) method* provides such a correction. The (IV) method is a two-

*The IV Method is discussed fully in Chapter 9: Henry Theil, *Principles of Econometrics*, Wiley & Sons, New York, 1971

stage technique. In the first stage, the variable subject to measurement error is regressed on a set of known correlated variables derived from other instruments. For example, MVMT scores, MFFT error rate, Parent Interview (child/family demographics) data might be regressed against PSI pre-test scores. A set of predicted values are generated from this first-stage regression and used as data for the mis-measured variable in the originally-posited regression relationship. The (IV) estimator is biased but consistent (the bias decreases as the sample size increases), whereas the ordinary regression estimator is biased and inconsistent as well.

Simultaneous equations bias arises when the connection between two variables in a regression analysis occurs in both directions simultaneously. Such interactions are most likely to arise during Phase II in the estimation of relationships in the cost model. For example, one equation in the cost model would explain the average wage paid to caregivers in a given day care center in terms of the professional qualifications held by caregivers in that center. Whether or not it is recognized explicitly in the cost model, there is clearly another relationship at work by which the professional qualifications of caregivers the center has been able to obtain is influenced by the wage rate the center pays. In effect, professional qualifications and wage rates are simultaneously (jointly) determined; the causation between them goes both ways. This joint causality can be shown to introduce bias into statistical results when the uncorrected regression formula is used for estimation. Since the relationship between cost and professionalism is a critical issue in the National Day Care Study, this bias must be reduced. The instrumental variable method described above for measurement bias can also be applied to the simultaneous equations problem. Again, the IV method will not eliminate the bias completely but should significantly reduce it for the sample sizes we will have available.

Chapter 6: A FRAMEWORK FOR PHASE III PLANNING

The Phase III design will be based on a combination of policy requirements and feasibility, as determined during Phase II. The approach to the development of the design is based on a systematic application of Phase II information, much as Phase I field and survey data were used to select Phase II sites and centers. The Phase III plan will be developed by a sequential set of decisions based on these data as described in this chapter, and the plan will be continually revised if new information or criteria warrant changes. This chapter briefly discusses the framework within which decision rules and actual Phase III decisions will be made during the spring of 1976.

The design for Phase III will include these components:

- Instrument selection
- Development of basic experimental design
- Selection of sites, centers, experimental manipulations
- Plan for experimental manipulations

This chapter assumes that several steps of Phase II research have been completed and the results are available:

- A list will have been developed of the independent variables that might be experimentally manipulated during Phase III of the study. Since only a few variables can actually be manipulated, this list will consist of a subset of the variables presented in Table 6-1. Staff/child ratio and at least one component of professionalism (educational attainment, previous experience, etc.) are certain to be chosen as design variables.
- An analysis relating process variables to independent variables will have been completed according to plans presented earlier in this volume. This analysis is to be completed in the T1.5 analytic period.

Table 6-1
Partial List of Independent Variables

- **Center Variables (Programmatic)**
  - Ratio
  - Group Size
  - Center Size
  - Age Mixing
  - Auspices, Funding
  - Age of Center
  - Services
  - Educational Program
- **Center Variables (Staff)**
  - Professionalism
    - Education
    - Experience
  - Staff Age/Race/Sex
  - Staff Turnover
- **Family Variables**
  - Child Age/Sex/Race
  - Family Income
  - Family Attitudes/Expectations
  - Family Structure
  - Child's Previous Day Care Experience

- An analysis involving both T1 and T2 scores will be available in time for final selection of Phase III measures, but not in time for selection of sites, centers, and a basic design.

The Phase III basic design will include the following components, each of which must be resolved:

- Choice of variables (1) to be experimentally manipulated and analyzed as predictors of outcome, (2) to be monitored

at constant values to the extent possible, (3) that may be allowed to vary, possibly confounded with other variables.

- Selection of a factorial design for basic replication, including choice of experimental levels and feasible experimental manipulations.
- Choice of site or sites and of replication patterns within and across sites.

These three design components are not independent and the choices available are jointly constrained by the parameters of present Phase II design and by the actual opportunities available for controls and manipulations. For this reason, the basic alternatives should be presented before the details of specific decision mechanisms are examined.

The RFP as well as previous proposals and reports agree that about 32 of the present 64 centers will be retained in one or more of the Phase II sties. Phase II has been designed to make possible such a decrease in study units by an analytic narrowing on the basis of Phase II results. We will also consider retaining more than 32 centers if a larger number would increase the power of study findings.

The following table presents four basic options in distributing approximately 32 study centers in one, two and three sites. The table is meant to depict broad options, not small details, so that, for example, option III might actually involve a 10/10/10 center distribution.

Exact numbers to be selected in each site depend on the exact number in a desired experimental replicate (identical experimental blocks) and the exact total number of study centers. Options with fewer than eight centers in one or more sites cannot be considered because they reduce experimental blocks to unacceptably small sizes. The first three options afford equal representation of centers in one, two, or

Table 6-2
Alternative Site Designs

Site Design Options	Number of Sites	Fractional Distribution of Centers	Example Center Distribution		
			Atlanta	Detroit	Seattle
I	1	All in Atlanta	32	–	–
II	2	One-half in each of two sites to be selected	16	–	16
III	3	One-third, one-third, one-third in each of present three sites	12	12	12
IV	3	One-half in Atlanta, one-fourth in Seattle, one-fourth in Detroit, proportional to present distribution	16	8	8

three sites, while the last retains the proportions of the Phase II distribution. Thus, the third option may suggest three 3 x 2 x 2 replicates, for a total of 36 centers, while the fourth would suggest retention of four 2 x 2 x 2 replicates and 32 centers, but either might be changed to fit a different factorial design.

The question of the feasibility and practicality of multi-level designs needs to be reviewed. The intent of the RFP was that staff/child ratio and professionalism, and perhaps other variables, should be set and maintained at discrete fixed levels, and that the basic analysis should be performed as ANOVA or ANCOVA. It will be necessary to deviate somewhat, but not substantially, from this plan for the following reasons:

- New levels of the variables, the staff/child ratio for instance, may not be set to fixed levels such as 1/5 or 1/15 very exactly without great expense and substantial intrusion. The manipulated variables will depend on accidents in the original circumstance as well as the manipulation. The extent of center reorganization should be minimized.
- Even if ratios were set to specified values, attrition or addition of teachers and children would change these values throughout Phase III. Observations taken to date show large fluctuations in enrollment. We will try to exercise control but such control cannot be perfect.
- Actual operating ratios also vary substantially throughout the day during different activity periods and as a result of different schedules of caregivers and children.

Since true fixed-level designs are not feasible, ordinary ANOVA or ANCOVA is not a feasible means of analysis. The variability of independent variables demands use of multiple regression models of the types discussed earlier. However, statistical power is much enhanced by ensuring that variables have considerable variance and are nearly independent of each other. Approximate implementation of a factorial design will lead to an efficient experimental design, even though it will not be analyzable by simple ANCOVA.

Since implementation of even a two-level factorial design will be very imperfect, and quantities can be expected to vary over the course of Phase III, it is not necessary or desirable to consider more than three levels for any variable. The principal reason that multi-level designs are used by ANOVA is to assure that data exist on which to base estimates of curvilinearity. The imperfect implementation of two-level designs is sufficient to assure this capability in Phase III, so that no explicit effort need be made to implement four or more levels in a design. Even three levels may be unnecessary to ensure estimability of curvilinearity. However, three-level designs improve flexibility in fitting designs totalling about 32 centers, and they will be considered, along with two-level designs, in Phase III. Table 6-3 exhibits the flexibility permitted under these constraints.

Table 6-3
Alternative Factorial Designs

Alternative Factorial Design	Number of Experimental Factors	Number of Centers Per Replicate
2 x 2	2	4
2 x 3	2	6
3 x 3	2	9
2 x 2 x 2	3	8
3 x 2 x 2	3	12
2 x 2 x 2 x 2	4	16
3 x 3 x 2	3	18
2 x 2 x 2 x 2 x 2	5	32
3 x 3 x 2 x 2	4	36

Table 6-4

Potential Site/Factorial Designs

(One or More Full Replicates Overall)

Site Design	Number of Sites	Factorial Design	Number of Centers Per Replicate	Number of Replicates Per Site	Total Study Centers
I	1	2 x 2	4	8	32
		2 x 3	4	6	36
		3 x 3	9	4	36
		2 x 2 x 2	8	4	32
		3 x 2 x 2	12	3	36
		2 x 2 x 2 x 2	16	2	32
		3 x 3 x 2	18	2	36
		3 x 3 x 2 x 2	36	1	36
		2 x 2 x 2 x 2 x 2	32	1	32
II	2	2 x 2	4	4/4	32
		2 x 3	6	3/3	36
		3 x 3	9	2/2	36
		2 x 2 x 2	8	2/2	32
		2 x 2 x 2 x 2	16	1/1	32
		3 x 3 x 2	18	1/1	36
		2 x 2 x 2 x 2 x 2	32	1/1	64
		2 x 2 x 2 x 2 x 2	32	½/½	32
III	3	2 x 3	6	2/2/2	36
	3	3 x 2 x 2	12	1/1/1	36
IV	3	2 x 2	4	4/2/2	32
		3 x 3	9	2/1/1	36
		2 x 2 x 2	8	2/1/1	32
		2 x 2 x 2*	8	4/2/2	64
		2 x 2 x 2 x 2	16	1/½/½	32
		2 x 2 x 2 x 2 x 2	32	½/¼/¼	32
		2 x 2 x 2 x 2 x 2	32	1/½/½	64

*Current Design

Combining the options of Table 6-2 with those of Table 6-3 leads to the design options exhibited in Table 6-4. Two 64-center designs have been included to provide additional alternatives. Site and center selection plans will not be greatly limited by constraining the options to those displayed. Designs involving fractional replication involve too many factors and constraints to hope for even approximate implementation.

The basic building block of each design is the definition of the replicate or block which is a factorial design in two or more variables.

For instance, the 2 x 2 design would include staff/child ratio and professionalism, each at two levels and fully crossed. The 3 x 2 x 2 design would treat staff/child ratio at three levels to obtain a more uniform spread, while retaining professionalism and group size at two levels. For a replicate of given size, there is a tradeoff between the levels of representation of policy variables and other independent variables, and in the choice of variables.

The second part of the design is the assignment of replicates to sites. If a potential site effect is not to be confounded with any other effect or interaction, then whole numbers of replicates would be assigned to each site. The Phase II design is an example in which a basic 2 x 2 x 2 factorial design is replicated four times in Atlanta and twice each in Seattle and Detroit, so that 64 centers are divided 32/16/16 among the sites.